Adaptive and efficient quantile estimation: From deconvolution to Lévy processes

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. rer. nat.

im Fach Mathematik

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von
Dipl.-Math. Mathias Trabs

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II
Prof. Dr. Elmar Kulke

Gutachter:
1. Markus Reiß
2. Alexander Lindner
3. Alexander Goldenshluger

Tag der Verteidigung: 17. Juni 2014

Abstract

The estimation of quantiles and realated functionals is studied in two inverse problems: the classical deconvolution model and the Lévy model, where a Lévy process is observed and where we aim for the estimation of functionals of the jump measure.

From a more abstract perspective we study semiparametric efficiency in the sense of Hájek–Le Cam for functional estimation in regular indirect models. A general convolution theorem is proved which applies to a large class of statistical inverse problems. In particular, we consider the deconvolution model, where we prove that our plug-in estimators of the distribution function and of the quantiles are efficient. In the nonlinear Lévy model based on low-frequent discrete observations of the Lévy process, we deduce an information bound for the estimation of functionals of the jump measure. The strong relationship between the Lévy model and the deconvolution model is given a precise meaning.

Quantile estimation in deconvolution problems is studied comprehensively. In particular, the more realistic setup of unknown error distributions is covered. Under minimal and natural conditions we show that the plug-in method is minimax optimal. A data-driven bandwidth choice yields optimal adaptive estimation. The concept of quantiles is generalized to the possibly infinite Lévy measures by considering left and right tail integrals. Based on equidistant discrete observations of the process, we construct a nonparametric estimator of the generalized quantiles and derive minimax convergence rates. As a motivating financial example for inverse problems, we empirically study the calibration of an exponential Lévy model for asset prices. The estimators of the generalized quantiles are adapted to this model. We construct an optimal adaptive quantile estimator and apply the procedure to real data of DAX-options.

Zusammenfassung

Die Schätzung von Quantilen und verwandten Funktionalen wird in zwei inversen Problemen behandelt: dem klassischen Dekonvolutionsmodell sowie dem Lévy-Modell in dem ein Lévy-Prozess beobachtet wird und Funktionale des Sprungmaßes geschätzt werden.

Im einem abstrakteren Rahmen wird semiparametrische Effizienz im Sinne von Hájek-Le Cam für Funktionalschätzung in regulären, inversen Modellen untersucht. Ein allgemeiner Faltungssatz wird bewiesen, der auf eine große Klasse von statistischen inversen Problem anwendbar ist. Im Dekonvolutionsmodell beweisen wir, dass die Plugin-Schätzer der Verteilungsfunktion und der Quantile effizient sind. Auf der Grundlage von niederfrequenten diskreten Beobachtungen des Lévy-Prozesses wird im nichtlinearen Lévy-Modell eine Informationsschranke für die Schätzung von Funktionalen des Sprungmaßes hergeleitet. Die enge Verbindung zwischen dem Dekonvolutionsmodell und dem Lévy-Modell wird präzise beschrieben.

Desweiteren wird die Quantilschätzung für Dekonvolutionsprobleme umfassend untersucht. Insbesondere wird der realistischere Fall von unbekannten Fehlerverteilungen behandelt. Wir zeigen unter minimalen und natürlichen Bedingungen, dass die Plugin-Methode minimax optimal ist. Eine datengetriebene Bandweitenwahl erlaubt eine optimale adaptive Schätzung. Indem wir auf positiven und negativen Strahlen integrieren, werden Quantile auf den Fall von Lévy-Maßen, die nicht notwendiger Weise endlich sind, verallgemeinert. Mittels äquidistanten, diskreten Beobachtungen des Prozesses werden nichtparametrische Schätzer der verallgemeinerten Quantile konstruiert und minimax optimale Konvergenzraten hergeleitet. Als motivierendes Beispiel von inversen Problemen untersuchen wir ein Finanzmodell empirisch, in dem ein Anlagengegenstand durch einen exponentiellen Lévy-Prozess dargestellt wird. Die Quantilschätzer werden auf dieses Modell übertragen und eine optimale adaptive Bandweitenwahl wird konstruiert. Die Schätzmethode wird schließlich auf reale Daten von DAX-Optionen angewendet.

Contents

1. Introduction

Quantile estimation is at the heart of statistical inference. For practitioners estimated quantiles are highly relevant, but they depend in a nonlinear way on the underlying law. For directly observed i.i.d. samples the empirical quantiles have many preferable properties, for instance, they have the minimal asymptotic variance. In more complex models quantile estimation may however not be obvious. This is especially the case if the target distribution is not directly observable but "hidden" by some operator, and thus we have to address a possibly ill-posed inverse problem. The driving example of an indirect model in this thesis is inference on the jump measure of a Lévy process which we observe at discrete time points. For convenience this model is denoted as Lévy model.

Before we present our theoretical contributions, we empirically study the calibration of an exponential Lévy model for asset prices as a motivating example for inverse problems. Afterwards, we aim for semiparametric efficiency in the sense of Hájek–Le Cam for the Lévy model, which turns out to be a surprisingly difficult question. To answer it, we consider the problem from the abstract perspective of general inverse problems. From that point of view we find that the Lévy model is closely related to the classical and simpler deconvolution model. Therefore, we first investigate the quantile estimation problem in deconvolution, which is a widely used model itself. The gained insights are finally applied to quantile estimation in the Lévy model, where we revisit the financial example from the beginning.

Statistical models

Nonparametric deconvolution models are of high practical importance and lead to challenging questions in statistical methodology. Let $X_1, \dots, X_n$ be independent random variables. Suppose that we merely observe the random variables

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n, \tag{1.1}$$

that is the original (X_j) corrupted by i.i.d. error variables ε_j, independent of (X_j). The main objective is to estimate the τ-quantile, for $\tau \in (0,1)$,

$$q_\tau := \inf \{x \in \mathbb{R} : P(X_1 \leqslant x) \geqslant \tau\}$$

of the population X_1 from the observations $Y_1, \dots, Y_n$. The first nonparametric analysis of this model goes back to Carroll and Hall (1988) and Fan (1991a,b). In this classical literature the distribution of the measurement error is assumed to be completely known which is usually not satisfied in practice. Instead, we may assume that we have at hand a sample from the error distribution given by

$$\varepsilon_1^*, \dots, \varepsilon_m^*, \qquad m \in \mathbb{N}.$$

A kernel density estimator for unknown error distributions was introduced by Neumann (1997) assuming independence of the samples (Y_j) and (ε_k^*). Motivated by applications, we will allow for dependencies between both samples. In particular, our procedure applies to the experimental setup of repeated measurements.

In the Lévy model we observe a real-valued Lévy process $L = \{L_t : t \geqslant 0\}$. Due to the Lévy–Itô decomposition, it can be represented as the sum of a deterministic drift determined by a parameter $\gamma \in \mathbb{R}$, a Brownian motion with volatility $\sigma^2 > 0$ and an independent jump component. The distribution of the jump sizes is described by the so-called *Lévy measure* ν. In the fundamental case of a compound Poisson process ν is a finite measure. In general, ν may be infinite, satisfying $\int (x^2 \wedge 1) \mathrm{d}\nu < \infty$, and thus the Lévy process can have infinite jump activity. Since there is a one-to-one correspondence between the Lévy process and the triplet (σ^2, γ, ν), the latter is called *characteristic triplet*. Lévy processes are one of the main building blocks for modeling stochastic processes in biology, finance and physics.

By definition $\nu(A)$, for any Borel set $A \in \mathscr{B}(\mathbb{R})$, is the expected number of jumps per unit time whose size belongs to A. Taking into account the possible singularity of ν at zero, we define the generalized distribution function

$$N(t) := \begin{cases} \nu((-\infty, t]), & \text{for } t < 0, \\ \nu([t, \infty)), & \text{for } t > 0. \end{cases}$$

Our aim is to estimate its inverse function which we call *generalized quantile function*. For a given level $\tau \in (0, \nu(\mathbb{R}_\pm))$ we introduce

$$q_\tau^- := \sup\{t \geqslant 0 : \nu((-\infty, -t]) \geqslant \tau\} \quad \text{and} \quad q_\tau^+ := \sup\{t \geqslant 0 : \nu([t, \infty)) \geqslant \tau\}.$$

Intuitively, q_τ^+ (resp. q_τ^-) is the smallest value such that the expected number of positive (resp. negative) jumps with absolute size larger than q_τ^+ (resp. q_τ^-) is less than τ. Alternatively, jumps larger than q_τ^+ are expected once in $1/\tau$ time units.

We will consider two different observation schemes: If we directly observe the Lévy process at equidistant discrete time points $\Delta, 2\Delta, \dots, n\Delta$, the estimator relies on the increments $L_{\Delta k} - L_{\Delta(k-1)}$ which are independent and identically distributed according to the law of L_Δ. We will focus on low-frequency observations where $\Delta > 0$ remains fixed while $n \to \infty$. Nonparametric estimation in this case was first studied by Neumann and Reiß (2009).

The second observation scheme is motivated by the application in finance that is precisely described in Chapter 2. Modeling an asset as exponential Lévy model driven by the process L, we use prices of put and call options to estimate the characteristics of the Lévy process under the risk-neutral measure. Since the observed option prices are noisy, we face a statistical problem of regression type. The error analysis of nonparametric estimators is then similar to the case of low frequent direct observations and was studied by Cont and Tankov (2004b) and Belomestny and Reiß (2006a). In this application the generalized quantiles are an interesting risk measure.

The deconvolution model and the Lévy model are structurally different. In particular, the deconvolution model is linear while the Lévy model is not. Nevertheless, it has been noted by Belomestny and Reiß (2006a), Kappus (2012) and others that they are closely related to each other. One aim of this work is to study this phenomenon in more detail. Analyzing the risk for estimators in these two models, we see that the estimation errors have very similar features. More precisely, the estimation error in the Lévy model is dominated by a deconvolution problem with the distribution of the Lévy process itself and thus may be described as *auto-deconvolution*. However, upper bounds, like convergence rates or asymptotic distributions, are mainly properties of the estimators. In contrast, lower bounds reveal the deeper information theoretic structure. Studying information bounds in both models, we find that both models are in fact equivalent in the weak sense of Le Cam (1972).

The plug-in approach

To estimate the quantiles, we will apply the following general strategy. First, we construct a kernel density estimator $\widehat{f}_h$ for some bandwidth $h > 0$. Applying a plug-in approach, we obtain an estimator for the distribution function

$$\widehat{F}_h(x) = \int_{\infty}^{x} \widehat{f}_h(x)\mathrm{d}x.$$

To define a quantile estimator, two natural strategies may be pursued. Either the distribution function estimator is inverted, which would require an additional monotonization of $\widehat{F}_h$, or an M-estimation paradigm is applied. Following the plug-in idea, we will apply the second possibility and define the minimum-contrast estimator of the quantile q_τ for $\tau \in (0,1)$ by

$$\widehat{q}_{\tau,h} := \mathrm{argmin}_\eta \, |\widehat{F}_h(\eta) - \tau| = \mathrm{argmin}_\eta \Big| \int_{-\infty}^{\eta} \widehat{f}_h(x)\mathrm{d}x - \tau \Big|.$$

Finally, we have to choose the bandwidth h appropriately. Deriving minimax convergence rates for $n \to \infty$, we determine the asymptotically optimal bandwidth. This oracle choice depends on the unknown properties of the underlying distributions and is consequently not available for the practitioner. Applying the approach by Lepski (1990), we construct a completely data-driven choice of the bandwidth which yields the optimal convergence rates up to logarithmic factors. Lepski's crucial insight was that with an appropriate estimator of the size of the stochastic error, in terms of confidence intervals, we can find in a finite number of bandwidths the one which mimics the oracle choice. In order to prove such a result, we need exponential concentration inequalities for our estimators.

Since our method estimates the three most important quantities of a population, namely density, distribution function and quantile function, based on the same estimator, the procedure is very attractive. Bickl and Ritov (2003) have pointed out that it is indeed desirable if a minimax optimal estimator $\widehat{f}_h$ can be "plugged-in" to estimate a functional, which allows for $\sqrt{n}$-consistent estimation, not only with the parametric rate but even semiparametrically efficient. That means in the nonparametric model, the estimator of the derived parameter converge with parametric rate to a normal distribution with minimal variance. Bickl and Ritov call this *plug-in property* of the estimator $\widehat{f}_h$. We will go one step further and adopt this concept to our ill-posed inverse problems: We say that $\widehat{f}_h$ has the plug-in property if (i) it estimates the density with the optimal nonparametric rate r_n, (ii) the plug-in estimators based on $\widehat{f}_h$ achieve the optimal nonparametric rate $r_n' < r_n$, (iii) if $r_n' = n^{-1/2}$ then the plug-in estimator is semiparametrically efficient.

This concept outlines the major program of this thesis. We establish minimax rates for density, distribution function and quantile estimation and we study semiparametric efficiency for the latter two quantities. In addition we consider variants of the Lepski-method to ensure applicability of the quantile estimators.

Literature and own contributions

While inverse problems appear in many different shapes in the literature, information bounds are studied only in a few linear cases. In the linear indirect regression model, semiparametric efficiency and convolution theorems have been studied by Klaassen et al. (2001, 2005) and Khoujmane et al. (2007). In linear indirect density estimation problems, van Rooij et al. (1999) have derived a convolution theorem for a restricted class of linear functionals. In Söhl and Trabs (2012b) we show efficiency in the deconvolution problem. Using the polar decomposition or specific properties of the operators, all these studies are restricted to linear models. However, in many situations the operator K might not be linear, see e.g. Engl et al. (1996) and Bissantz et al. (2004). Hence, new mathematical methods are necessary. To the best of the author's knowledge semiparametric efficiency for nonparametric nonlinear inverse problems, for example the Lévy model, is a completely open problem.

The following efficiency results are contained in Trabs (2013). In view of the equivalence results by Brown and Low (1996) and Nussbaum (1996), the prototype of an inverse problem is to estimate $\vartheta \in \Theta \subseteq X$, or derived parameters, from observations $y_{\varepsilon,\vartheta}$ in the white noise model

$$y_{\varepsilon,\vartheta} = K(\vartheta) + \varepsilon \dot{W} \quad \text{for a continuous operator } K : X \supseteq \Theta \to Y,$$

where X and Y are Hilbert spaces and $\varepsilon \dot{W}$ denotes white noise on Y with noise level $\varepsilon > 0$. Studying minimax convergence rates when K is linear, Goldenshluger and Pereverzev (2000, 2003) have shown that the parametric rate ε can be achieved for linear functionals of ϑ whose smoothness is not smaller than the ill-posedness of the operator K.

Inspired by the results from van der Vaart (1991), we restate the classical local asymptotic normality theory, which was initiated by Hájek (1970), in a way that is appropriate to capture the inverse structure of the above mentioned models. Here, the linear white noise model serves as the local limit experiment in the sense of Le Cam (1972). This leads to the notion of *regular indirect models*, meaning that the white noise model is the locally linear weak approximation of the statistical experiment. The asymptotic linear structure is described by the so called *generalized score operator*. We derive a version of the Hájek–Le Cam convolution theorem, see Theorem 3.8, for the estimation of derived parameters for regular inverse problems. Although we focus on linear functionals in the examples, the theory applies to any parameter which is differentiable in a pathwise sense which we demonstrate for quantile estimation.

We show that the white noise model with a (possibly) nonlinear operator, the deconvolution model as well as the Lévy model are regular indirect models and thus the convolution theorem applies. In many cases estimators are known that have the optimal limit distribution and consequently the information bound is sharp.

In the deconvolution setting, we can relax the assumptions on the functionals and the admissible error densities by van Rooij et al. (1999), as well as the conditions on the smoothness and decay behavior of the densities of X_1 and ε_1 by Söhl and Trabs (2012b) substantially. In fact, the abstract approach leads to natural assumptions in the explicit models. Considering a general class of linear functionals which can be estimated with the parametric rate, we show as in Söhl and Trabs (2012b) that the plug-in estimator satisfies a central limit theorem (in fact, a Donsker theorem is stated in Söhl and Trabs (2012b), but this will not be discussed in this thesis) restricting to known error distributions. If the deconvolution problem is sufficiently mildly ill-posed, this class of functionals contains the distribution function. Based on this result, asymptotic normality of the quantile estimator is derived. Together with the convolution theorem we conclude in Theorem 3.30 that the distribution function estimator and the quantile estimator in the deconvolution model are indeed efficient.

Our information bounds in the Lévy model (Corollary 3.34 and Theorem 3.40) are of special interest for three reasons: First, it is an important paradigm for nonlinear problems in indirect density estimation and a canonical probabilistic inverse problem. Second, we want to understand from an efficiency perspective the auto-deconvolution structure of the Lévy model. Third, we answer the conjecture by Nickl and Reiß (2012) that their distribution function estimator is efficient in the Hájek–Le Cam sense. Since Buchmann and Grübel (2003) have constructed for a finite and known jump activity a decompounding estimator with smaller asymptotic variance, an information bound is of particular interest. With the general convolution theorem at hand, we can prove that both estimators by Nickl and Reiß (2012) and by Buchmann and Grübel (2003) are indeed efficient and thus prior knowledge of the jump intensity simplifies the statistical problem significantly.

Although the literature on deconvolution problems is extensive and very broad, the problem of adaptive deconvolution with unknown measurement errors was addressed only recently, see Comte and Lacour (2011); Johannes and Schwarz (2013) and Kappus (2014) for adaptive density estimation with unknown error distributions in the model selection framework. Minimax results and other properties for non-adaptive methods are given by Neumann (1997, 2007), Meister (2004), Delaigle et al. (2008), Johannes (2009) among others. In the setting of known error distribution asymptotic normality of estimators for the distribution function has been shown by van Es and Uh (2005) and Hall and Lahiri (2008). The problem of quantile estimation in deconvolution was considered only by Hall and Lahiri (2008). They have constructed a quantile estimator for the case of known error distributions by inverting the distribution function estimator, without proposing an adaptive bandwidth choice. Further references on deconvolution may be found in the monograph by Meister (2009).

As we shall establish, the error of the quantile estimator is directly related to that of the distribution function estimator. Yet, the general analysis of the latter was not clear before. Fan (1991b) has proposed an estimator for the distribution function by integrating the density deconvolution estimator. In order to perform an exact analysis of its variance, a truncation of the integral was required in the estimation procedure.

This resulted in a non-optimal estimation method for the case of ordinary smooth errors and raised the conjecture that 'plug-in does not work optimally' for distribution function estimation in deconvolution. Trying to circumvent this problem, Hall and Lahiri (2008) as well as Dattner et al. (2011) have constructed a distribution function estimator based on a direct inversion formula. It remained an open and intriguing question whether the canonical plug-in estimator for distribution or quantile function estimation yields asymptotically optimal results. Following Dattner et al. (2013), we show that under suitable conditions the distribution function and the quantiles can be estimated with $\sqrt{n}$-rate by integrating the density estimator and applying the M-estimation approach, respectively. Moreover, Theorem 4.7 states general nonparametric rates if the error distribution is unknown and has to be estimated. Lower bounds show that the nonparametric rates are minimax optimal. The rates in combination with the previous efficiency result yield that the density estimator has the plug-in property as described above. A variant of the Lepski-method is established that achieves the minimax rates up to a logarithmic payment for adaptation (Theorem 4.12).

Relying on the insights from the deconvolution model, we then study the Lévy model which is in various aspects more involved. In the high-frequency regime, i.e. $\Delta \downarrow 0$, we almost see the jumps in the path and thus the estimation of the Lévy measure is relatively straight forward, see for instance Basawa and Brockwell (1982); Figueroa-López and Houdré (2006) or Aït-Sahalia and Jacod (2012) for a review. As noticed by Neumann and Reiß (2009) in the low-frequency regime, $\Delta > 0$ fixed, the estimation problem is more difficult because the number of jumps that occurred within an increment is not identifiable. Using however the Lévy–Khintchine formula which links the characteristic function of the marginal distributions and the characteristic triplet, we can estimate the jump measure by a spectral approach. This idea was initiated by Belomestny and Reiß (2006a) and then studied further, see Gugushvili (2009); Trabs (2014); Reiß (2013) and references therein. The question of an adaptive estimation method in the high-frequency and low-frequency regime has been addressed by Comte and Genon-Catalot (2011) and Kappus (2014), respectively, who apply a model selection procedure. Nickl and Reiß (2012) and Nickl et al. (2013) have derived uniform central limit theorems for the generalized distribution function of the Lévy measure in the low-frequency and high-frequency regime, respectively. Confidence statements have been studied by Figueroa-López (2011) and Söhl (2014).

With the notable exception of Belomestny (2010), who estimates the fractional oder of a Lévy process, all previously mentioned articles consider only linear functionals of the jump measure. Instead, we solve the substantially more demanding problem of estimating the nonlinear generalized quantiles in the nonlinear inverse problem. The conditions we impose one the models are very weak. We allow the whole spectrum of Lévy processes, reaching from diffusions to pure jump processes with unbounded variation and the combination of both. The quantile estimators are very robust in the sense that they do not depend on the drift and volatility parameters, which may be surprising. In both above described observation schemes we derive convergence rates for the distribution function and the quantile estimator (Theorems 5.9 and 5.13). In view of the literature the rates appear to be minimax optimal. In the regression-type Lévy model, we provide an adaptive method of Lepski- type that again achieves the optimal rates and the additional loss appears to be unavoidable. The performance of the estimation procedure is illustrated in simulations and a real data study based on options on the German DAX-index.

The deconvolution operator

One of the most important tools of the mathematical analysis in the whole thesis is the deconvolution operator. Let us assume in the deconvolution model (1.1) that the laws of X_1 and ε_1 have Lebesgue densities f and f_ε, respectively. Consequently, Y_1 is distributed according to $f_Y = f * f_\varepsilon$. Denoting the characteristic functions of Y_1 and ε_1 by φ_Y and φ_ε, respectively, we obtain the Fourier inversion formula

$$f = \mathcal{F}^{-1}\Big[\frac{\varphi_Y}{\varphi_\varepsilon}\Big] = \mathcal{F}^{-1}\Big[\frac{1}{\varphi_\varepsilon}\Big] * f_Y,$$

where the second equality has to be understood in distributional sense. Hence, the deconvolution operator is given by $g \mapsto \mathcal{F}^{-1}[\mathcal{F}g/\varphi_\varepsilon]$. Slightly abusing notation, we denote it by $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$. Replacing φ_Y by the empirical characteristic function of (Y_j) and regularizing with a band-limited kernel K and bandwidth $h > 0$, the previous display gives us immediately the natural kernel density estimator, which was proposed by Fan (1991b)

$$\widehat{f}_h := \mathcal{F}^{-1}\Big[\frac{\mathcal{F}K(h\bullet)}{\varphi_\varepsilon}\Big] * \mu_n$$

for the empirical measure $\mu_n = \sum_{j=1}^n \delta_{Y_j}$ with Dirac measure δ_y in the point $y \in \mathbb{R}$. To estimate the linear functional $\int \zeta(x)f(x)\mathrm{d}x$ for suitable functions ζ, say $\zeta \in L^2(\mathbb{R}) \cap L^\infty(\mathbb{R})$, the plug-in approach yields

$$\begin{aligned}\int \zeta(x)\widehat{f}_h(x)\mathrm{d}x &= \int \zeta(x)\Big(\mathcal{F}^{-1}\Big[\frac{\mathcal{F}K(h\bullet)}{\varphi_\varepsilon}\Big] * \mu_n\Big)(x)\mathrm{d}x\\ &= \int \Big(\mathcal{F}^{-1}\Big[\frac{\mathcal{F}K(-h\bullet)}{\varphi_\varepsilon(-\bullet)}\Big] * \zeta\Big)(x)\mu_n(\mathrm{d}x).\end{aligned}$$

Since the regularizing $\mathcal{F}K(h\bullet)$ degenerates to one as $h \downarrow 0$, we have to study the mapping properties of the deconvolution operator $\mathcal{F}^{-1}[1/\varphi_\varepsilon(-\bullet)]$. To this end the Fourier multiplier approach by Nickl and Reiß (2012) is extremely useful. Under certain assumptions they show that $1/\varphi_\varepsilon$ is a Fourier multiplier on Besov spaces. In the same spirit Schmidt–Hieber et al. (2013) discuss the behavior of the deconvolution operator as pseudo-differential operator. To apply abstract Fourier multiplier theorems, an appropriate decay behavior of φ_ε is required which leads to restrictions for the allowed error distribution in the deconvolution model.

For unknown error distributions the deconvolution operator $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$ is not observable and we have to study the estimated counterpart as a random Fourier multiplier. We prove that it preserves the mapping properties of the deterministic $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$, but its operator norm turns out to be (slightly) larger.

In the Lévy model the estimation relies on the Lévy-Khintchine representation of the characteristic exponent of the infinitely divisible marginals L_Δ of the Lévy process. The estimators are thus based on the logarithm of the empirical measure φ_n. Linearizing the estimation error, we obtain

$$\log \varphi_n - \log \varphi = \log \Big(\frac{\varphi_n}{\varphi}\Big) \approx \frac{\varphi_n - \varphi}{\varphi}$$

for the characteristic function φ of L_Δ and thus the deconvolution operator comes into play again. Let us stress that $1/\varphi$ is again unknown, but because it appears only in the error analysis we do not have to estimate it. In that specific aspect the Lévy model is simpler than the deconvolution model with unknown error distributions. We will show that the Fourier multiplier property of $\mathcal{F}^{-1}[1/\varphi]$ is very natural in the context of infinitely divisible distribution, see Theorem 5.5: A polynomial decay of the characteristic function is necessary and already sufficient to conclude that $1/\varphi$ is a Fourier multiplier on Besov spaces (modulo mild regularity conditions). Interestingly, in both models the generalized score operator in the convolution theorem is given by the deconvolution operator.

Organization of the thesis

In Chapter 2 we introduce the exponential Lévy model. As a motivating financial example for inverse problems involving Lévy processes, we nonparametrically calibrate models according to the procedures by Belomestny and Reiß (2006a) and Trabs (2014), observing prices of European put and call options. The numerically efficient implementation of the spectral estimation procedures is discussed for Lévy models of finite jump activity as well as for self–decomposable Lévy models. We compare the performance of the procedures for finite and infinite jump activity based on options on the German DAX index and find that both methods achieve good calibration results. The stability of the finite activity model is studied when the option prices are observed in a sequence of trading days. The findings in the chapter will be published in Söhl and Trabs (2014).

In Chapter 3 we study semiparametric efficiency in the sense of Hájek–Le Cam for functional estimation in regular indirect models. A convolution theorem is proved which applies to a large class of statistical inverse problems. This is illustrated for the prototypical white noise and deconvolution model. The abstract approach is especially useful for nonlinear models. We prove that the plug-in estimators of the distribution function and of the quantiles are efficient. The Lévy model is discussed in detail, concluding an information bound for the estimation of functionals of the jump measure. The strong relationship between the Lévy and the deconvolution model is given a precise meaning. Most of the results in this chapter can be found in Trabs (2013). The central limit theorem for the distribution function estimator in deconvolution is proved in Söhl and Trabs (2012b).

In Chapter 4 quantile estimation in deconvolution problems is studied comprehensively. In particular, the more realistic setup of unknown error distributions is covered. Under minimal and natural conditions we show that the plug-in method is minimax optimal. We find a data-driven bandwidth choice which yields optimal adaptive estimation. As a side result we conclude minimax rates for the plug-in estimation of distribution functions with unknown error distributions. The method is applied to simulations and a real data example of blood pressure measurements. All the results are contained in the paper Dattner et al. (2013).

Chapter 5 contains the estimation of the generalized quantiles in the Lévy model. Before, we show that under natural conditions on the Lévy process a Fourier multiplier theorem can be applied. Based on equidistant discrete observations of the process, we construct a nonparametric estimator of the generalized quantiles and derive convergence rates which seem to be optimal in the minimax sense. Moreover, the estimator is adapted to the exponential Lévy model from Chapter 2 where we discuss an adaptive estimation method. We apply the procedure to real data of DAX-options.

Acknowledgements

First of all I thank my supervisor Markus Reiß for his excellent mentoring and for sharing his fascination for mathematics with me. His challenging questions and the stimulating discussions have promoted this thesis enormously.

The instructive and fruitful collaborations with my coauthors are gratefully acknowledged. Due to his very precise and accurate way of working, Jakob Söhl reliably found every mathematical weakness and provided new solutions. Itai Dattner infected me with the fun he has in mathematics and the discussions with Richard Nickl showed me new perspectives for mathematical statistics.

I have profited a lot from the excellent research environment I have found at the Humboldt-Universität zu Berlin. Mathematical discussions as well as coffee breaks have been gained a lot from Randolf Altmeyer, Markus Bibinger, Moritz Jirak, Hilmar Mai, Nicolas Perkowski and especially David Prömel. Thanks to all of you.

This research was partially supported by the Deutsche Forschungsgemeinschaft DFG through the FOR 1735 “Structural Inference in Statistics” and the SFB 649 “Economic Risk”. Part of the work on quantile estimation for deconvolution was done during a visit to EURANDOM.

I thank my friends and my family for their support. This thesis would not have come into being without the constant encouragement and the great personal support from Nicole. The laughing of Fabian cheered me up at the final spurt for this thesis.

Notation

Although not always explicitly mentioned, we throughout work on sequences of statistical experiments $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ for some common parameter set Θ which is usually infinite dimensional. Convergence rates will be established in the $\mathcal{O}_P$-sense which corresponds to the loss function of confidence intervals. To study optimality in the minimax sense, we define the uniform stochastic Landau symbol $\mathcal{O}_{P,\Theta}$ over the parameter set Θ: For random variables $(A_n)_{n\in\mathbb{N}}$ we write $A_n = \mathcal{O}_{P,\Theta}(1)$ if

$$\lim_{R\to\infty} \limsup_{n\to\infty} \sup_{\vartheta\in\Theta} P_\vartheta(A_n > R) = 0. \tag{1.2}$$

All function spaces which will be used in the text are defined in Appendix A. Let us gather the frequently used symbols in the following list.

$\xrightarrow{P}$	stochastic convergence
$\Rightarrow$	weak convergence
$\int$, $\int_A$	integral of the real line and integral over a set A
$\overset{d}{=}$	equally distributed
$\lesssim$	smaller or equal than up to a constant which is independent of the parameters involved
$\gtrsim$	larger or equal than up to a constant which is independent of the parameters involved
$\ll$	absolute continuity
$\langle \bullet, \bullet \rangle$	scalar product (canonical scalar product of $\mathbb{R}^d$ if not further specified)
$\langle \alpha \rangle$	largest integer which is strictly smaller than $\alpha > 0$
$\mathbb{1}_A$	indicator function of a set A, defined by $\mathbb{1}_A(x) = 1$ if and only if $x \in A$
A^c	complement of A
$A^\perp$	orthogonal complement of A
$\overline{A}$	closure of A
$\mathbb{B}^\star$	dual space of Banach space $\mathbb{B}$
$\mathscr{B}(\mathbb{R})$	σ-algebra of Borel sets on the real line
$\mathcal{C}^\alpha, \mathcal{D}^\beta$	nonparametric parameter classes in Chapter 4, cf. (4.7), (4.8)
$\mathcal{C}^s, \mathcal{D}^s, \mathcal{E}^s, \widetilde{D}^{s,s'}_\tau, \widetilde{E}^{s,s'}_\tau$	nonparametric parameter classes in Chapter 5, cf. (5.7), (5.10)
δ_x	Dirac measure in $x \in \mathbb{R}$
$\mathcal{F} g$	Fourier transform of $g \in L^1(\mathbb{R}) \cup L^2(\mathbb{R})$ defined by $\mathcal{F} g(u) := \int e^{iux} g(x)\mathrm{d}x$, $u \in \mathbb{R}$.
$\mathcal{F} \mu$	Fourier transform of a measure μ defined by $\mathcal{F} \mu(u) := \int e^{iux} \mu(\mathrm{d}x)$, $u \in \mathbb{R}$.

$\Gamma(k, \eta)$	Gamma distribution with shape parameter k and scale parameter η
$K^\star$	adjoint of the operator K
$K^\dagger$	Moore–Penrose pseudo inverse of K, cf. (3.3)
$\operatorname{dom} K$	domain of the operator K
$\ker K$	kernel of the operator K
$\operatorname{ran} K$	rang of the operator K
$L^p, H^s, C^s, B^s_{p,q}, BV$	function spaces, cf. Appendix A
μ^{*n}	n-fold convolution of the measure μ
$\mathcal{N}(\mu, \Sigma)$	multivariate normal distribution
$\mathcal{O}(1), o(1)$	Landau notation
$\mathcal{O}_P(1), o_P(1)$	stochastic Landau notation
$\mathcal{O}_{P,\Theta}(1), o_{P,\Theta}(1)$	uniform stochastic Landau symbols over a class Θ, cf. (1.2)
$v^\top$	transpose of the vector $v \in \mathbb{R}^d$
$x-$	left limit defined by $x- := \lim_{h\downarrow} x - h$
$x+$	right limit defined by $x+ := \lim_{h\downarrow 0} x + h$

2. Inverse problems in finance: Option calibration

In recent years exponential Lévy models are frequently used as a basis for the purpose of pricing and hedging. To understand even more complex models better, it is thus essential to study the Lévy models in detail. Assuming a constant and known riskless interest rate $r \geq 0$ and an initial value $S_0 > 0$, these models describe the price of a stock by

$$S_t = S_0 e^{rt+L_t}, \quad t \geqslant 0, \tag{2.1}$$

where $L = \{L_t : t \geqslant 0\}$ is a Lévy process. Thus jumps of the price process are taken into account and heavy tails in the returns are modeled appropriately. It has been shown that exponential Lévy models are capable of reproducing not only the volatility smile but also the fact that it becomes more pronounced for shorter maturities, cf. Cont and Tankov (2004a). Hence, they are more adequate for recovering the stylized facts of financial time series than the classical model by Black and Scholes (1973). To apply model (2.1), for example, for derivative pricing, one has to infer the Lévy triplet (σ^2, γ, ν) under the risk–neutral measure from observable data, since the triplet determines completely the distributional properties of the stock S. Focusing on an applied perspective, we discuss the estimation of the characteristics based on a finite sample of vanilla option prices in this chapter.

From this starting point it is an open and demanding question whether the calibrated model can be used to estimate derived parameters, especially the generalized quantiles of the Lévy measure. In Chapter 5 we will give a positive answer: The estimated jump density can be plugged-in to obtain an optimal quantile estimator. Both problems, the model calibration and the quantile estimation, turn out to be ill-posed inverse problems leading to interesting theoretical and practical questions. This chapter focuses on an applied perspective. It is based on the paper by Söhl and Trabs (2014).

Exponential Lévy models are studied in a wide range of pricing problems, for instance by Asmussen et al. (2004); Cont and Voltchkova (2005); Ivanov (2007). The calibration has mainly focused on parametric models, cf. Barndorff-Nielsen (1998); Eberlein et al. (1998); Carr et al. (2002) and the references therein. First nonparametric calibration procedures for finite activity Lévy models were proposed by Cont and Tankov (2004b) as well as by Belomestny and Reiß (2006a). That means in these approaches no (finite dimensional) parametrization is assumed. The method of Trabs (2014) extends the spectral calibration to the infinite activity case, more precisely to self–decomposable Lévy processes.

The calibration of a completely general Lévy process might be too much to hope for. We consider two submodels, precisely stated in Section 2.2 together with the constuction of the estimators. Under the first setup, denoted by (FA), the process L is assumed to be a jump–diffusion whose Lévy measure ν has finite total mass. In the second case, which we refer to as (SD), we consider a self–decomposable Lévy process without diffusion component, that is $\sigma = 0$. In particular, in the second setting the Lévy process has infinite jump activity and thus the two setups are non–overlapping. In both cases we do not assume that the Lévy density ν belongs to some parametric class. The explicit estimators for (σ^2, γ, ν) in the two models (FA) and (SD) are constructed essentially as in Belomestny and Reiß (2006a) and Trabs (2014), respectively, but some modifications are introduced which improve their numerical performance.

As shown in simulations in Section 2.3 these improvements reduce the mean squared error of the estimators significantly. In contrast to the method by Cont and Tankov (2004b) the spectral calibration is a straightforward algorithm, where no minimization problem has to be solved. Therefore, the methods are quite fast owing to the Fast Fourier transform (FFT). Whereas the above mentioned works focus on the asymptotic theory, we concentrate on the application of the method to realistic sample sizes in this chapter. In a related framework of a jump–diffusion Libor model, Belomestny and Schoenmakers (2011) study the application of the spectral calibration method to finite sample data sets. For a more precise inference we use confidence intervals as constructed by Söhl and Trabs (2014), relying on the results by Söhl (2014).

In Section 2.4 we use data of vanilla options on the German DAX index to compare the finite activity model to the self–decomposable one. Considering options with different maturities, both models achieve good calibration results in the sense that the residuals between the given data and the calibrated model are small. Since the Blumenthal–Getoor index equals zero in our models, the calibration based on option data behaves quite differently from the case of high–frequency observations under the historical measure, where Aït-Sahalia and Jacod (2009) find evidence that the Blumenthal–Getoor index is larger than one. Applying the calibration to a sequence of trading days, we obtain the evolution of the model parameters in time. The estimators seem to be stable with respect to the spot time.

2.1. Lévy processes

Let us first recall some basic properties of Lévy processes which can be found in Sato (1999).

Definition 2.1. The $\mathbb{R}^d$-valued stochastic process $L = \{L_t : t \geqslant 0\}$ defined on a probability space $(\Omega, \mathscr{F}, P)$ is a *Lévy process* if it satisfies:

(i) L is stochastically continuous, i.e., for any $t \geqslant 0$, $L_u \xrightarrow{P} L_t$ as $u \to t$,

(ii) $P(L_0 = 0) = 1$,

(iii) For any $n \in \mathbb{N}$ and $0 \leqslant t_0 < \cdots < t_n$ the random variables $L_{t_0}, L_{t_1} - L_{t_0}, \dots, L_{t_n} - L_{t_{n-1}}$ are independent *(independent increments)*,

(iv) For any $0 \leqslant s < t$ it holds $L_t - L_s \overset{d}{=} L_{t-s}$ *(stationary increments)*.

Any Lévy process has a modification which is right-continuous with left limits. The marginal distributions of a Lévy process are infinitely divisible in the following sense:

Definition 2.2. A probability measure μ on $\mathbb{R}^d$ is infinitely divisible if for any $n \in \mathbb{N}$ there is a probability measure μ_n on $\mathbb{R}^d$ such that $\mu = \mu_n^{*n}$.

In fact, there is a one-to-one correspondence between Lévy process and infinitely divisible distributions. The most important result on Lévy processes for this thesis is the Lévy–Khintchine representation.

Proposition 2.3 (Lévy–Khintchine formula)**.** *Let L be an $\mathbb{R}^d$-valued Lévy process. For any $t \geqslant 0$ the characteristic function of L_t is given by*

$$\varphi_t(u) := \mathbb{E}[e^{i\langle u, L_t\rangle}] = e^{t\psi(u)}, \quad u \in \mathbb{R}^d,$$

with characteristic exponent

$$\psi(u) = -\frac{1}{2}\langle u, \sigma^2 u\rangle + i\langle \gamma, u\rangle + \int_{\mathbb{R}^d} (e^{i\langle x,u\rangle} - 1 - i\langle x, u\rangle \mathbb{1}_{|x|\leqslant 1})\nu(\mathrm{d}x) \qquad (2.2)$$

for a symmetric nonnegative-definite matrix $\sigma^2 \in \mathbb{R}^{d\times d}$, a drift parameter $\gamma \in \mathbb{R}^d$ and a Lévy measure ν on $\mathbb{R}^d$, i.e., ν is a σ-finite measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ with

$$\nu(\{0\}) = 0 \quad \textit{and} \quad \int_{\mathbb{R}^d} (|x|^2 \wedge 1)\nu(\mathrm{d}x) < \infty.$$

Since there is a bijection between the Lévy process L and the triplet (σ^2, γ, ν), the latter is called *characteristic triplet* or *Lévy triplet*. Throughout this theses we will consider only real valued processes. There are two special cases of the Lévy–Khintchine formula which are quite useful. If $d = 1$ and $\int(|x| \wedge 1)\nu(\mathrm{d}x) < \infty$, then ψ in (2.2) simplifies to

$$\psi(u) = -\frac{\sigma^2}{2}u^2 + i\gamma_0 u + \int (e^{iux} - 1)\nu(x)\mathrm{d}x, \quad u \in \mathbb{R}, \qquad (2.3)$$

for $\gamma_0 = \gamma - \int_{-1}^{1} x\nu(\mathrm{d}x)$. Assuming alternatively $d = 1$ and $\int x^2\nu(\mathrm{d}x) < \infty$, we can write ψ in Kolmogorovo's version

$$\psi(u) := -\frac{\sigma^2}{2}u^2 + i\gamma_1 u + \int (e^{iux} - 1 - iux)\nu(\mathrm{d}x), \quad u \in \mathbb{R},$$

with $\gamma_1 = \gamma - \int_{|x|>1} x\nu(\mathrm{d}x)$.

2.2. Spectral calibration

Let L be a real-valued Lévy process with characteristic triplet (σ^2, γ, ν). Throughout this chapter, we assume

(A1) ν is absolutely continuous. Abusing notation, we denote its Lebesgue density likewise by $\nu\colon \mathbb{R} \to \mathbb{R}_+$.

(A2) $\int(|x| \wedge 1)\nu(x)\mathrm{d}x < \infty$.

Owing to (A2), the jump component J_t has finite variation. Calibrating the exponential Lévy model (2.1) reduces to estimating the two one-dimensional parameters σ^2 and $\gamma_0 = \gamma - \int_{-1}^{1} x\nu(\mathrm{d}x)$ as well as the density ν from an infinite dimensional parameter space. However, the characteristic triplet depends on the underlying measure. Since we are interested in pricing and hedging purposes, we consider throughout the risk neutral measure under which the discounted process e^{L_t} is a martingale. Therefore, $\mathbb{E}[e^{L_t}] = 1, t \geq 0$, which is equivalent to the martingale condition

$$\frac{\sigma^2}{2} + \gamma_0 + \int_{-\infty}^{\infty} (e^x - 1)\nu(\mathrm{d}x) = 0. \tag{2.4}$$

This condition implicitly implies an exponential moment assumption on L_t and thus on ν, too.

So far, nonparametric calibration methods exist in two different setups:

(FA) Assumptions (A1) holds and Assumption (A2) is replaced by the stronger assumption of finite activity $\lambda := \int \nu(x)\mathrm{d}x < \infty$ (Cont and Tankov, 2004b; Belomestny and Reiß, 2006a; Söhl, 2014).

(SD) L_t is self–decomposable with $\sigma = 0$ that is ν can be characterized by $\nu(\mathrm{d}x) = k(x)/|x|\mathrm{d}x$ for $x \in \mathbb{R}\setminus\{0\}$, where k is increasing on $\mathbb{R}_-$ and decreasing on $\mathbb{R}_+$. Additionally, $\alpha := k(0+) + k(0-)$ is assumed to be finite (Trabs, 2014).

Note that in the (SD) setting Assumptions (A1) and (A2) are automatically satisfied. The function k with the above monotonicity properties is called k-function. Trabs (2014) considers a more general class of Lévy processes where k does not need to fulfill these monotonicity properties. However, we will see that the class of self–decomposable processes is already rich enough to calibrate the model (2.1) well.

Typical parametric submodels of (FA) and (SD) are given by Examples 2.4 and 2.5, respectively. We will use them to study the performance of estimation methods in simulations.

Example 2.4 (Merton model)**.** Merton (1976) has introduced the first exponential Lévy model. Therein, the jumps are normally distributed with intensity $\lambda > 0$:

$$\nu(x) = \frac{\lambda}{\sqrt{2\pi}v} \exp\left(-\frac{(x-\eta)^2}{2v^2}\right), \quad x \in \mathbb{R}.$$

A realistic choice of the parameters is $\eta = -0.1$, $v = 0.2$ and $\lambda = 5$. Together with the volatility $\sigma = 0.1$ this determines the drift to be $\gamma_0 = 0.379$ using the martingale condition (2.4).

Example 2.5 (Variance gamma model)**.** Let (W_t) be a standard Brownian motion and (G_t) an independent Gamma process with mean rate one and variance rate ρ that is $G_t \sim \Gamma(t/\rho, 1/\rho)$. Madan and Seneta (1990) defined the variance gamma process with parameters σ, ρ and ϑ as the time changed Brownian motion with drift $L_t = \vartheta G_t + \sigma W_{G_t}, t \geq 0$. This is a model with infinite jump activity and Blumenthal–Getoor index zero. The characteristic function and the k–function of L are given by

$$\varphi_t(u) = (1 + i\vartheta\rho u + \sigma^2\rho u^2/2)^{-t/\rho} \quad \text{and}$$
$$k_{VG}(x) = \frac{1}{\rho}e^{x/\eta_m}\mathbf{1}_{\{x<0\}}(x) + \frac{1}{\rho}e^{-x/\eta_p}\mathbf{1}_{\{x\geq 0\}}(x), \quad u, x \in \mathbb{R},$$

with $\eta_p := \sqrt{\vartheta^2\rho^2/4 + \sigma^2\rho/2} + \vartheta\rho/2$ and $\eta_m := \sqrt{\vartheta^2\rho^2/4 + \sigma^2\rho/2} - \vartheta\rho/2$, respectively. In our simulations we use the parameters $\sigma = 1.2, \rho = 0.2$ and $\vartheta = -0.15$. The value of $\gamma_0 = 0.141$ is given by the martingale condition again. These choices imply $\alpha = k_{VG}(0+) + k_{VG}(0-) = 10$.

Since we want to estimate the model parameters under the risk neutral measure, the procedure is based on observed prices of vanilla options. Throughout, we measure the time in years. Let us fix a maturity $T > 0$, define the negative log–moneyness $x := \log(K/S_0) - rT$ and denote call and put prices by $\mathcal{C}(x,T) = S_0\mathbb{E}[(e^{L_T} - e^x)_+]$ and $\mathcal{P}(x,T) = S_0\mathbb{E}[(e^x - e^{L_T})_+]$, respectively. In terms of the option function

$$O(x) := \begin{cases} S_0^{-1}\mathcal{C}(x,T), & x \geqslant 0, \\ S_0^{-1}\mathcal{P}(x,T), & x < 0, \end{cases} \tag{2.5}$$

our observations are given by

$$O_j = O(x_j) + \delta_j\varepsilon_j, \quad j = 1, \ldots, n, \tag{2.6}$$

with noise levels $\delta_j > 0$ and independent, centered errors ε_j, satisfying $\operatorname{Var}(\varepsilon_j) = 1$ as well as $\sup_j \mathbb{E}[\varepsilon_j^4] < \infty$. The observation errors are due to the bid–ask spread and other market frictions. For simplicity, we assume $(\delta_j)_{j=1,\ldots,n}$ to be known. Otherwise, the noise levels can be estimated on an independent data set, for instance, from market data which contain separately bid and ask prices. Note that since the Lévy density ν is an infinite–dimensional object, the triplet $(\sigma^2, \gamma_0, \nu)$ cannot be inferred from the market price of just one vanilla option as the volatility parameter in the Black–Scholes model. The more prices O_j are observed for different strikes x_j, the more accurate the estimation will be. To construct the estimators of the Lévy triplet, we apply the Lévy–Khintchine representation (2.3) and the pricing formula by Carr and Madan (1999)

$$\mathcal{F}O(u) := \int_{-\infty}^{\infty} e^{iux}O(x)\mathrm{d}x = \frac{1 - \varphi_T(u-i)}{u(u-i)}. \tag{2.7}$$

Note that the latter equation extends to all complex numbers in $\{u \in \mathbb{C} | \operatorname{Im}(u) \in [0,1]\}$ since there the characteristic function $\varphi_T(u-i)$ is finite by the exponential moment of L_T, which is implied by the martingale condition (2.4). We obtain

$$\psi(u) = \frac{1}{T}\log(1 - u(u+i)\mathcal{F}O(u+i)), \tag{2.8}$$
$$\psi_{-i}(u) := \psi(u-i) = \frac{1}{T}\log(1 + iu(1+iu)\mathcal{F}O(u)). \tag{2.9}$$

Through curve fitting to $(x_j, O_j)_{j=1,\ldots,n}$, we obtain an empirical versions $\widetilde{O}$ of the option function and subsequently, through a plug–in approach, empirical versions $\widetilde{\psi}$ and $\widetilde{\psi}_{-i}$ of the characteristic exponents. While the theoretical results (Belomestny and Reiß, 2006a; Trabs, 2014) concentrate on a linear interpolation of the observation, an additional smoothing by using B–splines of degree two might improve the estimators. In Section 2.3 we provide simulations with both interpolation methods to investigate the practical influence.

Given $\widetilde{\psi}$, we can estimate the characteristics of the process from the spectral representation. The procedures of Belomestny and Reiß (2006a) as well as Trabs (2014) rely on the identity (2.9) which looks more convenient because it uses directly the option function. The identity (2.8) uses an exponentially scaled option function since $\mathcal{F}O(u+i) = \mathcal{F}[e^{-x}O(x)](u)$. However, in (2.9) the characteristic exponent is shifted by $-i$, which leads to estimators of exponentially scaled versions of the jump density ν and of the k-function k, respectively. Therefore, we will use it only to estimate the one–dimensional parameters of the models. According to the idea of Belomestny and Reiß (2006b), equation (2.8) allows to estimate immediately the nonparametric objects ν and k. Regularization of the procedure is achieved by cutting off frequencies larger than a regularization parameter $U > 0$. Since (FA) and (SD) need to be considered separately, the precise estimators are given in Sections 2.2.1 and 2.2.2. Note that in both cases correction steps are necessary to satisfy non–negativity of the jump density and the martingale condition (2.4) (see Söhl and Trabs, 2012a, for details). If the latter one would be violated, the right–hand side of the pricing formula (2.7) could have a singularity at zero and thus we could not apply the inverse Fourier transform to obtain an option function from the calibration.

A critical question is the choice of the regularization parameter U. As a benchmark, we use in simulations an oracle cut–off value, that is U minimizes the discrepancy between the estimators and the true values of σ^2, γ_0 and ν measured in an L^2–loss. To calibrate real data, we employ the simple least squares approach

$$U^* := \operatorname{arginf}_{U>0} RSS(U) \tag{2.10}$$

with the residual sum of squares

$$RSS(U) := \sum_{j=1}^{n} |\widehat{O}_U(x_j) - O_j|^2,$$

where $\widehat{O}_U$ is the option function corresponding to the Lévy triplet estimated by means of the cut–off value U. We determine $\widehat{O}_U$ by the pricing formula (2.7) and Lévy-Khintchine representation (2.3), in which we plug in the estimators obtained by using the cut–off value U. The estimated option function $\widehat{O}_U$ can be computed efficiently for each U so that the numerical effort of finding U^* is mainly determined by the minimization algorithm used to solve (2.10). From theoretical consideration a penalty term, as used by Belomestny and Reiß (2006b), is necessary to avoid an over–fitting, that is not to choose U too large. Nevertheless, our practical experience with this method shows that the above mentioned correction steps, which are not included in the theory, lead to an auto–penalization: Using large cut–off values, the stochastic error in the estimators becomes large. This leads to high fluctuations of the nonparametric part and the correction has an increasing effect. Hence, the difference between $\widetilde{O}$ and $\widehat{O}_U$ becomes larger if U is too high and the residual sum of squares increases, too. In particular, imposing the jump density to be nonnegative implies a shape constraint on the state price density which is basically the second derivative of the option function. Therefore, the least squares choice of the tuning parameter works well at least for small noise levels.

The approach to minimize the calibration error was also applied by Belomestny and Schoenmakers (2011). Alternative data–driven choices of the cut–off value U are the "quasi-optimality" approach which was studied by Bauer and Reiß (2008) and which was applied by Belomestny (2011) or the use of a preestimator as proposed by Trabs (2014). However, we will consider only the least squares approach which performs well in our application.

2.2.1. The finite activity case

In the (FA) setup we deduce from (2.3) and (2.9) the identity

$$\psi_{-i}(u) = -\frac{\sigma^2}{2}u^2 + i(\sigma^2+\gamma_0)u + (\sigma^2/2+\gamma_0-\lambda) + \mathcal{F}\mu(u) \quad \text{with } \mu(x) := e^x\nu(x).$$

The estimators of the parameters are defined by Belomestny and Reiß (2006a) as follows:

$$\widehat{\sigma}^2 := \int_{-U}^{U} \mathrm{Re}(\widetilde{\psi}_{-i}(u))w_\sigma^U(u)\mathrm{d}u, \tag{2.11}$$

$$\widehat{\gamma}_{fa} := -\widehat{\sigma}^2 + \int_{-U}^{U} \mathrm{Im}(\widetilde{\psi}_{-i}(u))w_{\gamma_{fa}}^U(u)\mathrm{d}u, \tag{2.12}$$

$$\widehat{\lambda} := \frac{\widehat{\sigma}^2}{2} + \widehat{\gamma}_{fa} - \int_{-U}^{U} \mathrm{Re}(\widetilde{\psi}_{-i}(u))w_\lambda^U(u)\mathrm{d}u, \tag{2.13}$$

where w_σ^U, $w_{\gamma_{fa}}^U$ and w_λ^U are suitable weight functions and $\widetilde{\psi}_{-i}$ is the empirical version of ψ_{-i}. To avoid ambiguity, the estimator of γ_0 has an additional subscript denoting the model in which the estimator is defined. The estimators in (2.11),(2.12) and (2.13) can be understood as weighted L^2–projections of $\widetilde{\psi}_{-i}$ onto the space of quadratic polynomials. In this sense the estimators arise naturally as a solution of a weighted least squares problem. However, the optimal weight depends not only on the unknown heteroscedacity in the frequency domain but also on the unknown function $\mathcal{F}\mu$, so we do not pursue this approach here. Instead we construct the weight functions w_σ^U, $w_{\gamma_{fa}}^U$ and w_λ^U directly as Belomestny and Reiß (2006b) but propose different weight functions. The idea is that the noise is particularly high in the high frequencies and thus it is desirable to assign less weight to the high frequencies. A smooth transition of the weight functions to zero at the cut–off value improves the numerical results significantly. Therefore, we would like the weight function and its first two derivatives to be zero at the cut–off value. With the side conditions on the weight functions this leads to the following polynomials:

$$\begin{aligned} w_\sigma^U(u) := \frac{c_\sigma}{U^3}\Big((2s+1)\left(\frac{u}{U}\right)^{2s} - 4(2s+3)\left(\frac{u}{U}\right)^{2s+2} + 6(2s+5)\left(\frac{u}{U}\right)^{2s+4} \\ - 4(2s+7)\left(\frac{u}{U}\right)^{2s+6} + (2s+9)\left(\frac{u}{U}\right)^{2s+8}\Big)\mathbb{1}_{[-U,U]}(u), \\ w_{\gamma_{fa}}^U(u) := \frac{c_{\gamma_{fa}}}{U^2}\left(\left(\frac{u}{U}\right)^{2s+1} - 3\left(\frac{u}{U}\right)^{2s+3} + 3\left(\frac{u}{U}\right)^{2s+5} - \left(\frac{u}{U}\right)^{2s+7}\right)\mathbb{1}_{[-U,U]}(u), \end{aligned}$$

$$w_\lambda^U(u) := \frac{c_\lambda}{U}\Big((2s+3)\left(\frac{u}{U}\right)^{2s} - 4(2s+5)\left(\frac{u}{U}\right)^{2s+2} + 6(2s+7)\left(\frac{u}{U}\right)^{2s+4}$$
$$- 4(2s+9)\left(\frac{u}{U}\right)^{2s+6} + (2s+11)\left(\frac{u}{U}\right)^{2s+8}\Big)\mathbb{1}_{[-U,U]}(u),$$

where all three functions equal zero outside $[-U,U]$. The constants $c_\sigma, c_{\gamma_{fa}}, c_\lambda \in \mathbb{R}$ are determined by the normalization conditions

$$\int_{-U}^{U} u^2 w_\sigma^U(u)\mathrm{d}u = -2, \quad \int_{-U}^{U} u w_{\gamma_{fa}}^U(u)\mathrm{d}u = 1 \quad \text{and} \quad \int_{-U}^{U} w_\lambda^U(u)\mathrm{d}u = 1.$$

The parameter s reflects the a priori knowledge about the smoothness of ν. As a rule of thumb it can be chosen equal to two. The gain of the new weight functions is discussed in Section 2.3.

To estimate directly the jump density ν and not only the exponential scaled version μ, we use ψ instead of ψ_{-i} as discussed above. Therefore, we define the estimator

$$\widehat{\nu}(x) := \mathcal{F}^{-1}\Big[\Big(\widetilde{\psi}(u) + \frac{\widehat{\sigma}^2}{2}u^2 - i\widehat{\gamma}_{fa}u + \widehat{\lambda}\Big) w_\nu^U(u)\Big](x), \tag{2.14}$$

where $\widetilde{\psi}$ is the empirical version of ψ and w_ν^U is a flat top kernel with support $[-U,U]$:

$$w_\nu^U(u) := F\left(\frac{u}{U}\right) \quad \text{with} \quad F(u) := \begin{cases} 1, & |u| \le 0.05, \\ \exp\left(\frac{-\exp(-(|u|-0.05)^{-2})}{(|u|-1)^2}\right), & 0.05 < |u| < 1, \\ 0, & |u| \ge 1. \end{cases} \tag{2.15}$$

To evaluate the integrals in (2.11) to (2.13), it suffices to apply the trapezoidal rule. The inverse Fourier transformation in (2.14) can be efficiently computed using the FFT–algorithm. Therefore, depending on the interpolation method which is applied to obtain $\widetilde{O}$, the whole estimation procedure is very fast. Finally, we note that the cut–off value can be chosen differently for each quantity $\sigma^2, \gamma_0, \lambda$ and ν. A documentation of the implementation in R can be found in Söhl and Trabs (2012a).

2.2.2. The self–decomposable framework

Recall that $\sigma = 0$ is assumed in the (SD) setting. While the Blumenthal–Getoor index is zero in this case the parameter α describes the degree of activity of the process on a finer scale. To calibrate the self–decomposable model, we need a different representation of ψ_{-i} than before because of the infinite activity of these processes. Applying Fubini's theorem to (2.3), we obtain $\psi_{-i}(u) = i\gamma_0 u + \gamma_0 + \int_0^1 i(u-i)\,\mathcal{F}[\mathrm{sign}(x)k(x)]((u-i)t)\mathrm{d}t, u \in \mathbb{R}$, where the Fourier transform decays slowly since $\mathrm{sign}(x)k(x)$ is not continuous at zero. Trabs (2014, Prop. 2.2) showed that decomposing $\mathrm{sign}(x)k(x)$ into a nonsmooth and a smooth part yields for $u \neq 0$

$$\psi_{-i}(u) = D(u) + i\gamma_0 u - \alpha\log(|u|) + \sum_{j=1}^{2s-2}\frac{i^j(j-1)!\alpha_j}{u^j} + \rho(u), \tag{2.16}$$

where $2s$ is the smoothness of k away from zero, $\alpha_j := k^{(j)}(0+) + k^{(j)}(0-)$ for $j = 1, \dots, 2s-2$, the function D is constant on the real half lines and the remainder ρ corresponds to the smooth part of $\operatorname{sign}(x)k(x)$ and thus satisfies $\|u^{2s-1}\rho(u)\|_\infty < \infty$. Owing to the polynomial decay of ρ, estimators of γ_0 and α can be defined analogously to Section 2.2.1, filtering the coefficient of the linear term and of the logarithmic term in (2.16), respectively:

$$\begin{aligned}\widehat{\gamma}_{sd} &:= \int_{-U}^{U} \operatorname{Im}(\widetilde{\psi}_{-i}(u)) w^U_{\gamma_{sd}}(u)\mathrm{d}u,\\ \widehat{\alpha} &:= \int_{-U}^{U} \operatorname{Re}(\widetilde{\psi}_{-i}(u)) w^U_{\alpha}(u)\mathrm{d}u\end{aligned}$$

with polynomial weight functions

$$w^U_{\gamma_{sd}}(u) = \frac{1}{U^2}\sum_{k=0}^{s+1} a_k \Big(\frac{u}{U}\Big)^{2(k+s)-1} \quad \text{and} \quad w^U_{\alpha}(u) = \frac{1}{U}\sum_{k=0}^{s+1} b_k \Big(\frac{u}{U}\Big)^{2(k+s)},$$

where the coefficients $a_k, b_k \in \mathbb{R}$ are determined by

$$\begin{aligned}&\int_0^U u w^U_{\gamma_{sd}}(u)\mathrm{d}u = \frac{1}{2}, && \int_0^U w^U_{\gamma_{sd}}(u)\mathrm{d}u = 0 \quad \text{and} && \int_0^U u^{-2l+1} w^U_{\gamma_{sd}}(u)\mathrm{d}u = 0,\\ &\int_0^U \log(|u|) w^U_{\alpha}(u)\mathrm{d}u = -\frac{1}{2}, && \int_0^U w^U_{\alpha}(u)\mathrm{d}u = 0 \quad \text{and} && \int_0^U u^{-2l} w^U_{\alpha}(u)\mathrm{d}u = 0,\end{aligned}$$

for $l = 1, \dots, s-1$. These integral conditions lead to a system of linear equations which can be solved analytically as well as numerically. To estimate the k-function, we combine the method of Trabs (2014) with the approach by Belomestny and Reiß (2006b). From (2.3), Fubini's theorem and (2.8) follows

$$\psi'(u) = i\gamma_0 + i\mathcal{F}[\operatorname{sign}\cdot k](u) = \frac{(u - iu^2)\,\mathcal{F}[xO](u+i) - (2u+i)\,\mathcal{F}\,O(u+i)}{T(1 - u(u+i)\,\mathcal{F}\,O(u+i))}. \tag{2.17}$$

Let $\widetilde{\psi}'$ be the empirical version of ψ' obtained by substituting O by $\widetilde{O}$ in (2.17). Since we know the position of the jump of k, the application of a one–side kernel function allows to estimate the k–function on the whole real line. We define

$$\widehat{k}(x) := \begin{cases}\mathcal{F}^{-1}\big[(-\widehat{\gamma}_{sd} - i\widetilde{\psi}'(u))\mathcal{F}W_k(u/U)\big](x), & x > 0,\\ \mathcal{F}^{-1}\big[(\widehat{\gamma}_{sd} + i\widetilde{\psi}'(u))\mathcal{F}W_k(-u/U)\big](x), & x < 0,\end{cases}$$

with a one–sided kernel function

$$W_k(x) := \Big(\sum_{m=0}^{2s} c_m x^m\Big) F(x+1), \quad x \in \mathbb{R},$$

where F is the flat top kernel defined in (2.15), thus $\operatorname{supp} W_k = [-2, 0]$, and the coefficients $c_k \in \mathbb{R}$ are chosen such that

$$\int W_k = 1, \quad \int x^l W_k(x)\mathrm{d}x = 0 \quad \text{for } l = 1, \dots, 2s-1.$$

Again, the coefficients are given by a system of linear equations, which can be solved numerically. Because the kernel W_k has to be one–sided, it cannot have compact support in the Fourier domain. Hence, to ensure that there are no large stochastic errors in $\widetilde{\psi}'(u)$ for large $u \in \mathbb{R}$, a truncation in the spirit of Trabs (2014) might be reasonable. To obtain an estimator in the class of self–decomposable processes, we have to ensure the necessary monotonicity of the k–function. Therefore, we apply rearrangement, which is a general procedure to transform a function into a monotone function. With some arbitrary large constant $C > 0$, the rearranged estimator is then given by

$$\widehat{k}^*(x) := \begin{cases} \inf\Big\{y \in \mathbb{R}_+ \Big| \int_0^C \mathbf{1}_{\{\widehat{k}(z) \geqslant y\}} \mathrm{d}z \leqslant x\Big\}, & x \in (0, C], \\ \inf\Big\{y \in \mathbb{R}_+ \Big| \int_0^C \mathbf{1}_{\{\widehat{k}(-z) \geqslant y\}} \mathrm{d}z \leqslant |x|\Big\}, & x \in [-C, 0), \\ 0, & \text{otherwise.} \end{cases} \tag{2.18}$$

In the sequel, we identify $\widehat{k}$ with its rearranged version $\widehat{k}^*$, since we are interested only in the calibration using self-decomposable processes.

2.3. Simulations

Let us first describe the setting of all of our simulations. In view of the higher concentration of European options at the money, the design points $\{x_1, \dots, x_n\}$ are chosen to be the $k/(n+1)$–quantiles, $k = 1, \dots, n$, of a normal distribution with mean zero and variance $1/2$. The observations O_j are computed from the characteristic function φ_T using the fast Fourier transform. The additive noise consists of independent, normal and centered random variables with variance $\delta_j^2 = |t\mathcal{O}(x_j)|^2$ for some relative noise level $t > 0$. By choosing the sample size n and the deviation parameter t, we determine the noise level of the observations. According to the existing theoretical results, it is well measured by the quantity

$$\varepsilon := \Delta^{3/2} + \Delta^{1/2}\|\delta\|_{\ell^\infty} \quad \text{with} \quad \Delta := \max_{j=2,\dots,n}(x_j - x_{j-1}),$$

which takes the interpolation error and the stochastic error into account. The interest rate and time to maturity are set to $r = 0.06$ and $T = 0.25$, respectively.

Using the Merton model with the parameters of Example 2.4, we start with investigating the practical influence of two aspects of the procedure, which are mentioned above. The interpolation of the data (x_j, O_j) with linear B–splines is compared to the use of quadratic B–splines. The latter preprocessing is an additional smoothing of the data, which achieves significant gains for higher noise levels. The other point of interest is the choice of the weight functions. Since it is known from the theory that the noise affects mainly the high frequencies, the polynomial weight functions greatly reduce the variance of the estimator. These improvements are illustrated in Figure 2.1: In the case of $\widehat{\sigma}$ we approximate the root mean squared error (RMSE) $\sqrt{\mathbb{E}[|\widehat{\sigma} - \sigma|^2]}$ using 500 Monte–Carlo iterations with and without quadratic splines and polynomial weight functions, respectively. This is done for different noise levels, whereby t decreases from 0.03 to 0.015 and n increases from 50 to 400, simultaneously.

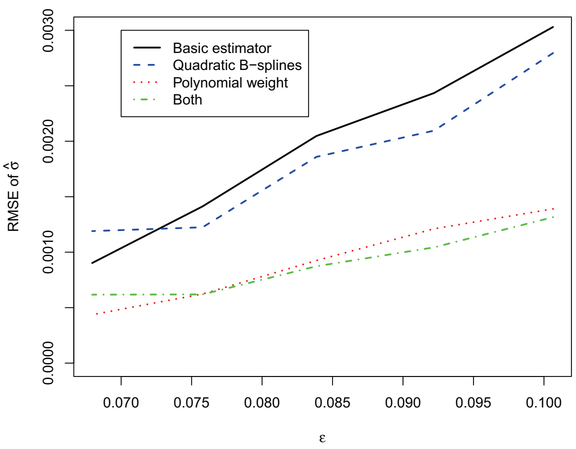

Figure 2.1.: RMSE of $\widehat{\sigma}$ for different noise levels with 500 Monte–Carlo iterations in each case. Usage of the linear and quadratic spline interpolation as well as usage of the weight functions by Belomestny and Reiß (2006a) and the polynomial weight functions.

To illustrate the performance of the estimation methods in the (FA) and the (SD) model, we will simulate $n = 100$ strike prices in the Merton model and in the variance gamma model with parameters as in Examples 2.4 and 2.5, respectively. The relative noise level is taken to be $t = 0.01$. To coincide with the theory, we interpolate the corresponding European call prices linearly. In the real data application in Section 2.4 we will take advantage of the B-spline interpolation.

To quantify the estimation error, we will additionally give in the sequel confidence statements for the above point estimators. Söhl (2014) shows asymptotic normality of the estimators in the (FA) setup. Based on these results but applying a finite sample point of view, Söhl and Trabs (2014) have constructed confidence intervals in both models, the finite activity case and the self–decomposable one. We will use their method without going into details on the construction.

Figure 2.2 shows the true Lévy density and k-function in the (FA) and the (SD) setting, respectively, their estimators with oracle choice of the cut-off values and the corresponding pointwise 95% confidence intervals. Almost everywhere the true function is contained in the confidence intervals. Moreover, another 100 estimators from further Monte Carlo iterations are plotted. The graphs show that the confidence intervals describe well the deviation of the estimated jump densities. The negative bias around zero might come from the smoothing which naturally tends to smooth out peaks, cf. Härdle (1990, Chap. 5.3).

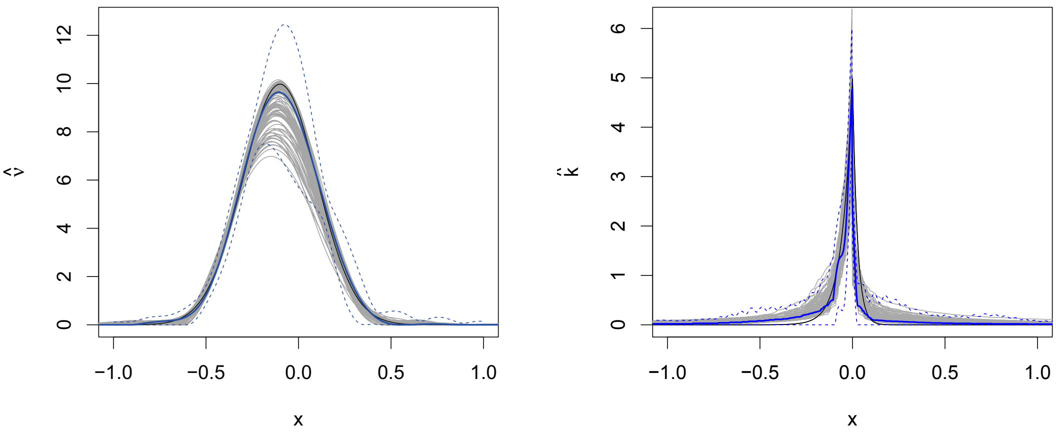

Figure 2.2.: True *(black, solid)* and estimated *(blue, bold)* Lévy density *(left)* and k-function *(right)* from a simulation of the Merton model and the variance gamma model, respectively. Pointwise 95% confidence intervals *(blue, dashed)* use the oracle cut–off values $U = 19$ and $U = 2.8$, respectively. Additional 100 estimators *(grey)* are shown from a Monte Carlo simulation.

2.4. Empirical study

We apply the calibration methods to a data set from the Deutsche Börse database Eurex[1]. It consists of settlement prices of European put and call options on the DAX index from May 2008. Therefore, the prices are observed before the latest financial crises and thus the market activity is stable. The interest rate r is chosen for each maturity separately according to the put-call parity at the respective strike prices. The expiry months of the options are between July and December, 2008, and thus the time to maturity T, measured in years, reaches from two to seven months. The number of our observations n is given in Figure 2.3 and lays around 50 to 100 different strikes for each maturity and trading day.

2.4.1. Comparison of (FA) and (SD)

Let us first focus on one (arbitrarily chosen) day. Hence, we calibrate the option prices of May 29, 2008, with all four different maturities to both, the (FA) and the (SD) setting. The results are summarized in Table 2.1 and Figure 2.4. Using the complete estimation of the models, we generate the corresponding option functions $\widehat{O}$. They are graphically compared to the given data points and we calculate the residual sum of squares $RSS = RSS(U^*)$ as defined in (2.10). For all maturities both methods yield good fits to the data. However, for longer maturities, especially the calibration of options with seven months to maturity, minor problems occur in the (SD) calibration. Although

[1]provided through the SFB 649 "Economic Risk"

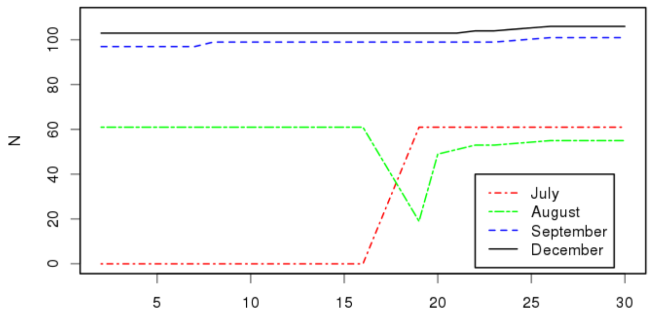

Figure 2.3.: Number of observed prices of put and call options during May, 2008.

the sample size is larger, the estimated standard deviation is larger for longer maturities in the (SD) scenario, too. The calibration at other trading days confirms this weakness of the (SD) method for larger T. This coincides with the asymptotic analysis of Trabs (2014) where longer durations lead to slower convergence rates of the risk.

Moreover, Figure 2.4 shows that the estimated option function $\widehat{O}$ which results from the (SD) calibration does not exactly recover the tails of O. In all maturities and in both models the Lévy density has more weight on the negative half line and thus there are more negative jumps than positive ones priced into the options. This coincides with the empirical findings in the literature (see eg, Cont and Tankov, 2004a). Due to the positivity correction, the jump densities might look unsmooth where they are close to zero. This problem might be circumvented by adding smoothness constraints. However, the construction of confidence intervals would then be much more difficult. Hence, this topic is left open for further research.

In view of the parametric calibration of their CGMY model Carr et al. (2002) suggested that risk-neutral processes of stocks should be modeled by pure jump processes with finite variation. Now, the nonparametric approach shows that both models the finite activity case and the self-decomposable model are able to reproduce the option data. The finite activity jump-diffusion seem to work even more robust with respect to T. Note that in both models the Blumenthal–Getoor index equals to zero which is in contrast to the investigation of high-frequency historical data, where Aït-Sahalia and Jacod (2009) estimated a jump activity index larger than one.

2.4.2. (FA) across trading days

The aim of this section is twofold. By considering more than one day we investigate the stability of the (FA) estimation procedure. Moreover, calibrating the model across the trading days in May, 2008, shows the development of the model along the time line and with small changes in the maturities. To profit from the higher observation number, we apply the calibration procedure for the (FA) case to the options with maturity in September and December.

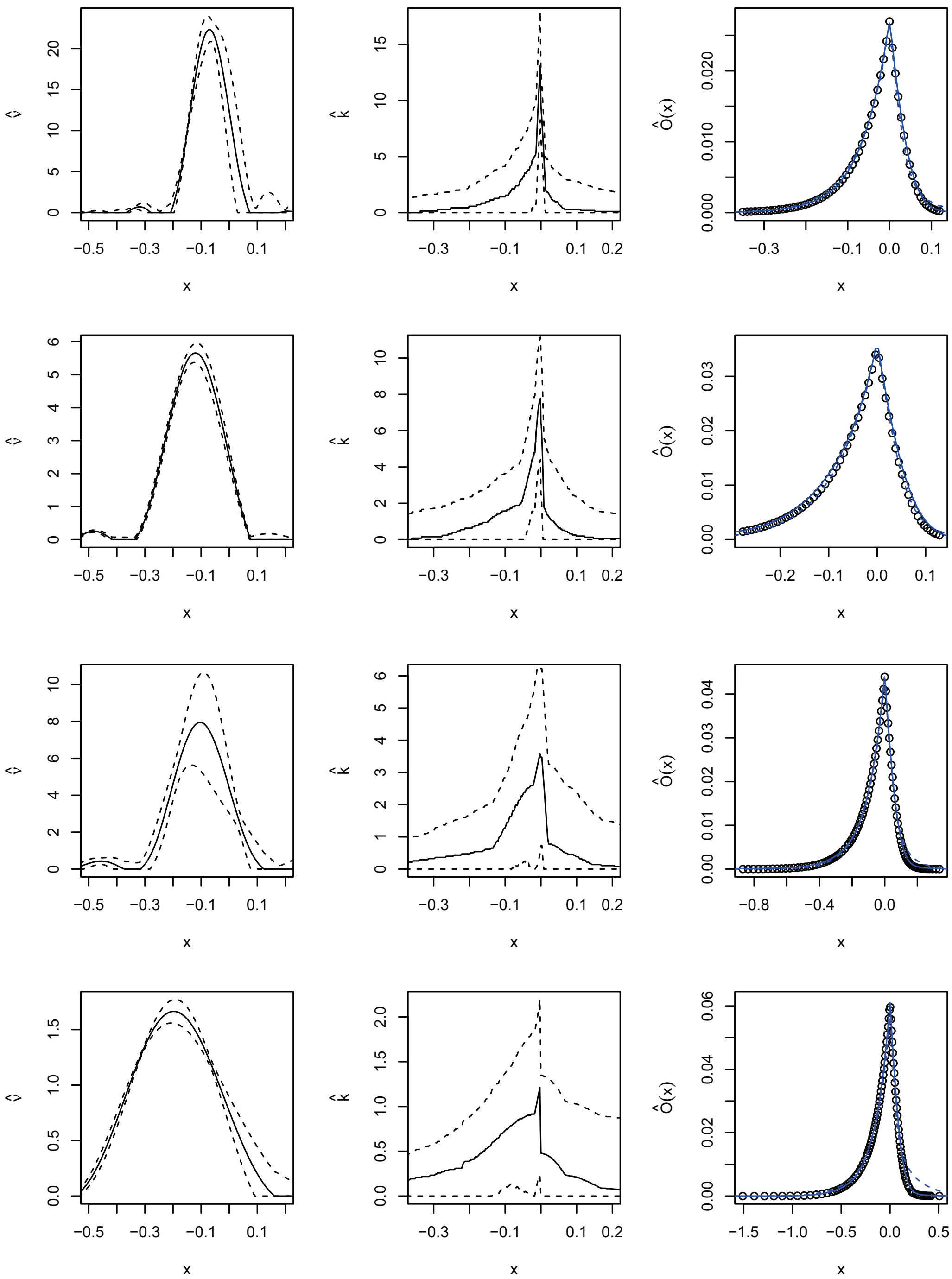

Figure 2.4.: Estimated jump densities *(left)*, k–functions *(center)* with pointwise 95% confidence intervals as well as calibrated option functions in the (FA) *(right, solid)* and (SD) *(right, dashed)* setting and given data from May 29, 2008 *(right, points)*. The time to maturity increases from $T = 0.136$ *(top)* to $T = 564$ *(bottom)*.

n	61		55		101		106	
T	0.136		0.233		0.311		0.564	
(FA)								
$\widehat{\sigma}$	0.110	(0.0021)	0.123	(0.0009)	0.107	(0.0030)	0.124	(0.0013)
$\widehat{\gamma}_{fa}$	0.221	(0.0049)	0.142	(0.0015)	0.174	(0.0050)	0.105	(0.0011)
$\widehat{\lambda}$	3.392	(0.2015)	1.290	(0.0176)	1.823	(0.1261)	0.637	(0.0181)
$\sqrt{RSS}$	0.003		0.008		0.005		0.008	
(SD)								
$\widehat{\gamma}_{sd}$	0.344	(0.0103)	0.336	(0.0136)	0.302	(0.3242)	0.139	(0.0607)
$\widehat{\alpha}$	8.662	(0.1534)	8.677	(0.2938)	3.670	(0.0797)	5.181	(1.0030)
$\sqrt{RSS}$	0.007		0.006		0.011		0.029	

Table 2.1.: Estimated parameters ϑ and estimated standard deviation $\widehat{s}_\vartheta$ (in brackets) for $\vartheta \in \{\sigma, \gamma_{fa}, \lambda, \gamma_{sd}, \alpha\}$ and residual sum of squares using option prices from May 29, 2008, with n observed strikes for each maturity T.

The estimations of the parameters are displayed in Figure 2.5. Note that we do not smooth over time. Furthermore, the 95% confidence intervals for the December options are shown. The estimated volatility $\widehat{\sigma}$ fluctuates around 0.1 and 0.12. The confidence sets imply that there is no significant difference between the two maturities. Both $\widehat{\gamma}_{fa}$ and $\widehat{\lambda}$ decrease for higher durations: On the one hand the curves of December lay significantly below the ones of September, on the other hand the graphs have a slight positive trend with respect to the time axis, which means with smaller time to maturity. Keeping in mind that the implied volatility in the Black–Scholes model typically decreases for longer time to maturity, this lower market activity is reproduced by smaller jump activities in our calibration while the volatility is relatively stable.

Figure 2.6 displays the estimated jump densities. All jump measures have a similar shape which is in line with real data calibration of Belomestny and Reiß (2006b). In contrast to Cont and Tankov (2004b) the densities are unimodal or have only minor additional modes in the tails, which may be artefacts of the spectral calibration method. The tails of $\widehat{\nu}$ do not differ significantly, while the different heights reflect the development of the jump activities $\widehat{\lambda}$. There is an obvious trend to small negative jumps in all data sets, which is in line with the stylized facts of option pricing models. The calibration is stable for consecutive market days.

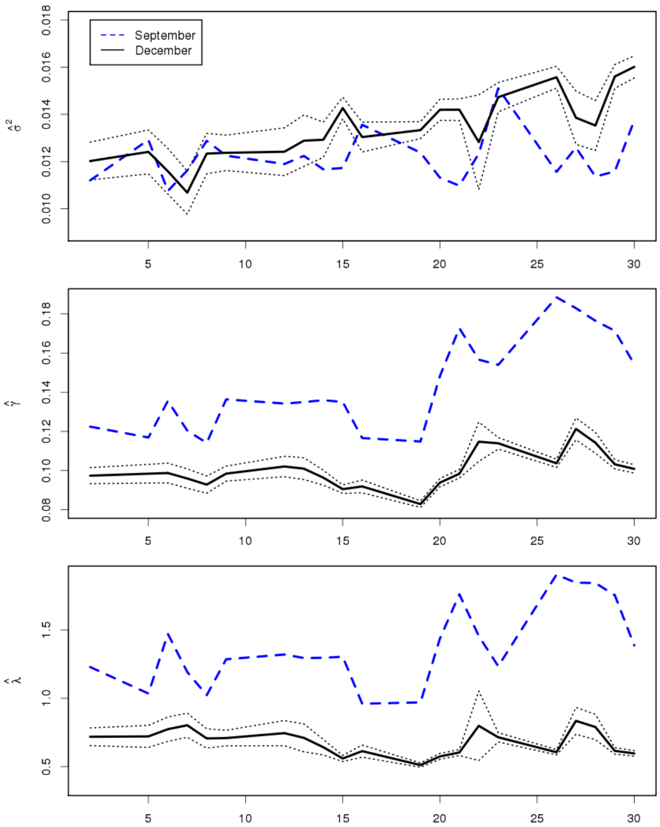

Figure 2.5.: At each market day in May, 2008, estimated σ^2 *(top)*, γ_0 *(center)* and λ *(bottom)* from options with maturities in September *(dashed)* and December *(solid)* and confidence intervals *(dotted)* for the latter ones.

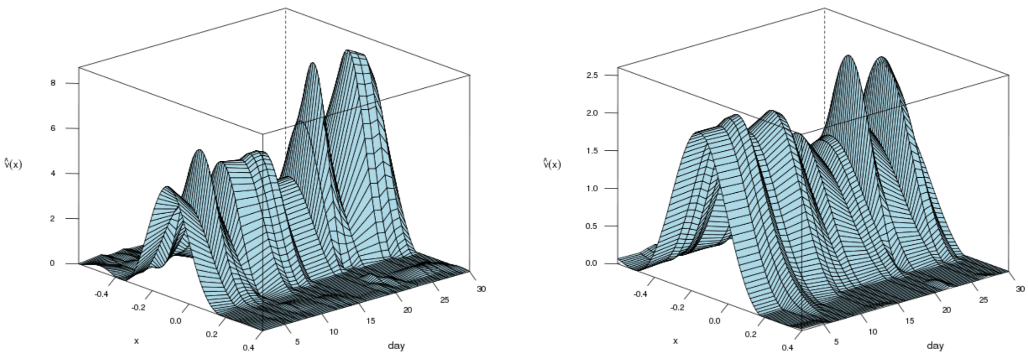

Figure 2.6.: Estimation of ν for maturity in September *(left)* and December (*right)*.

3. Semiparametric efficiency

Inverse problems have attracted enormous attention in applied mathematics in the last decades, in particular models with noise in the data, since typically the parameter which is the target of the statistical inference is not directly observable, but “hidden” by some operator. While upper bounds, like convergence rates, are mainly properties of the estimators, lower bounds reveal the deeper information theoretic structure.

Instead of the (infinite dimensional) parameter itself, derived quantities are often the final object of interest. On the hand, they might allow for inference with parametric rate, circumventing typical problems in nonparametric estimation like the choice of the bandwidth, cf. estimating the distribution function instead of the density. On the other hand, many nonparametric statistical procedures rely on basis expansions and model selection strategies, see e.g. Cavalier et al. (2002). For these adaptive methods it is strictly necessary to assess the quality of the estimated coefficient in terms of confidence. Consequently, information bounds are of particular interest.

In Section 3.1 we develop our general results. We start with efficient estimation in the linear white noise model

$$y_{\varepsilon,\vartheta} = K(\vartheta) + \varepsilon\dot{W} \quad \text{for a continuous operator } K : X \supseteq \Theta \to Y, \tag{3.1}$$

where X and Y are Hilbert spaces and $\varepsilon\dot{W}$ denotes white noise on Y with noise level $\varepsilon > 0$. Afterward, we introduce *regular indirect models* and the *generalized score operator*. We derive a version of the Hájek–Le Cam convolution theorem for the estimation of derived parameters for regular inverse problems. The tangent set is determined by the range of generalized score operator and the efficient influence function is given by the Moore–Penrose pseudoinverse of the adjoint score operator. We discuss the extension from $\mathbb{R}^d$-valued functionals to general Banach spaced valued parameters.

In Section 3.2 we apply the abstract convolution theorem to the deconvolution model. We provide the central limit theorem for the plug-in estimator for linear functionals and show that this estimator is semiparametrically efficient. From that result we deduce efficiency of the quantile estimator.

The general theory will be applied to the Lévy model in Section 3.3. To show the regularity of the Lévy model, we apply estimates of the distance of infinitely divisible distributions by Liese (1987). The resulting score operator has deconvolution structure. Surprisingly, its adjoint operator looks exactly the same as in the deconvolution case. The information bound coincides with the asymptotic distribution of the distribution function estimator by Nickl and Reiß (2012). More technical proofs are postponed to Section 3.4.

3.1. Information bounds for inverse problems

3.1.1. Linear white noise model

To understand the probabilistic structure of general inverse problems, we start with studying the abstract linear white noise model (3.1), where X and Y are separable real Hilbert spaces with scalar products $\langle \bullet, \bullet \rangle_X$ and $\langle \bullet, \bullet \rangle_Y$, respectively, and $K : X \to Y$ is a linear and bounded operator. To avoid identifiability problems, we additionally assume that K is injective. Therefore, we observe for some unknown $\vartheta \in X$

$$\langle y_{\varepsilon,\vartheta}, \varphi \rangle_Y = \langle K\vartheta, \varphi \rangle_Y + \varepsilon \dot{W}(\varphi) \quad \text{for all} \quad \varphi \in Y,$$

where $(\dot{W}(\varphi))_{\varphi \in Y}$ is an iso-normal Gaussian process with mean zero and covariance structure $\mathbb{E}[\dot{W}(\varphi_1)\dot{W}(\varphi_2)] = \langle \varphi_1, \varphi_2 \rangle_Y$ for $\varphi_1, \varphi_2 \in Y$. The law μ of the white noise $\dot{W}$ is defined as symmetric (zero mean) Gaussian measure on $(E, \mathscr{B}(E))$ for a separable Banach space E in which Y can be continuously embedded and where $\mathscr{B}(E)$ denotes the Borel σ-algebra on E. In other words $\dot{W}$ is an isometry from Y into $L^2(E, \mathscr{B}(E), \mu)$. For the construction of the so called abstract Wiener space we refer to Kuo (1975, Thm. 4.1, Lem. 4.7). We denote the law of $y_{\varepsilon,\vartheta}$ by $P_{\varepsilon,\vartheta}$.

Basically, the linear white noise model is a Gaussian shift experiment where the parameter is hidden by the operator K. The inverse problem is to estimate a derived parameter $\chi(\vartheta)$ from the observation $y_{\varepsilon,\vartheta}$ when $\varepsilon \to 0$. First, let us focus on a linear functional $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X$ for some $\zeta \in X$. Typically, K admits no continuous inverse, leading to an ill-posed problem, cf. Goldenshluger and Pereverzev (2000, 2003) or Cavalier (2008) for a recent review of nonparametric estimation.

Following the classical semiparametric approach, we study parametric submodels by perturbing the parameter ϑ in directions $b \in X$. For any $b \in X$ we consider the submodel $t \mapsto P_{\varepsilon,\vartheta_t}$ generated by the path $[0,1) \ni t \to \vartheta_t := \vartheta + tb$. The behavior of the submodel along this path is described by the following lemma.

Lemma 3.1. *Let $P_{\varepsilon,x}$ denote the law of $y_{\varepsilon,x} = K(x) + \varepsilon\dot{W}$ on $(E, \mathscr{B}(E))$ for $x \in X$ and an operator $K : X \to Y$, then for all $\vartheta \in X$*

$$\frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) = \exp\Big(\dot{W}\Big(\frac{K(x) - K(\vartheta)}{\varepsilon}\Big) - \frac{1}{2\varepsilon^2}\|K(x) - K(\vartheta)\|_Y^2\Big) \quad P_{\varepsilon,\vartheta} - a.s.$$

Proof. As discussed above the law of the white noise $\varepsilon\dot{W} = y_{\varepsilon,0}$ is a symmetric (zero mean) Gaussian measure on $(E, \mathscr{B}(E))$. Its (unique) reproducing kernel Hilbert space is Y with norm $\|\bullet\|_\varepsilon := \varepsilon^{-1}\|\bullet\|_Y$. To see this, note that every functional $\varphi \in E^\star$ can be represented by $\varphi = \langle y, \bullet \rangle_Y = \langle \widetilde{y}, \bullet \rangle_\varepsilon$ for some $y \in Y$ and $\widetilde{y} = \varepsilon^2 y$. Then

$$\varphi(\varepsilon\dot{W}) \sim \mathcal{N}(0, \|\widetilde{y}\|_\varepsilon^2) = \mathcal{N}(0, \varepsilon^2\|y\|_Y^2).$$

The Cameron–Martin formula for Gaussian measures on Banach spaces (Da Prato and Zabczyk, 1992, Prop. 2.24) yields that $P_{\varepsilon,x}$ and $P_{\varepsilon,0}$ are equivalent measures on $(E, \mathscr{B}(E))$ with Radon–Nikodym derivative

$$\frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,0}}(y_{\varepsilon,0}) = \exp\Big(\langle y_{\varepsilon,0}, K(x) \rangle_\varepsilon - \frac{1}{2}\|K(x)\|_\varepsilon^2\Big) \quad P_{\varepsilon,0} - a.s.$$

and thus

$$\begin{aligned}\frac{\mathrm{d}P_{\varepsilon,0}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) &= \exp\Big(- \langle y_{\varepsilon,\vartheta}, K(\vartheta)\rangle_\varepsilon + \frac{1}{2}\|K(\vartheta)\|_\varepsilon^2\Big) \quad \text{and}\\ \frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) &= \frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,0}}(y_{\varepsilon,\vartheta})\frac{\mathrm{d}P_{\varepsilon,0}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta})\\ &= \exp\Big(\langle \varepsilon\dot W, K(x)-K(\vartheta)\rangle_\varepsilon - \frac{1}{2}\|K(x)-K(\vartheta)\|_\varepsilon^2\Big)\\ &= \exp\Big(\Big\langle \dot W, \frac{K(x)-K(\vartheta)}{\varepsilon}\Big\rangle_Y - \frac{1}{2}\Big\|\frac{K(x)-K(\vartheta)}{\varepsilon}\Big\|_Y^2\Big) \quad P_{\varepsilon,\vartheta}-a.s. \qquad \square\end{aligned}$$

Linearity of K yields $\varepsilon^{-1}(K(\vartheta_\varepsilon)-K(\vartheta)) = Kb$ and thus by Lemma 3.1

$$\log\frac{\mathrm{d}P_{\varepsilon,\vartheta_\varepsilon}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) = \dot W(Kb) - \frac{1}{2}\|Kb\|_Y^2 \quad P_{\varepsilon,\vartheta}-a.s. \tag{3.2}$$

and therefore model (3.1) with linear operator K satisfies the classical LAN condition (even without local and asymptotic) with parameter $h = Kb \in \operatorname{ran} K$. To find an information bound for the derived parameter $\chi(\vartheta) = \langle\zeta,\vartheta\rangle_X$, we express it in terms of $\eta = K\vartheta$ by

$$\psi(\eta) := \langle \zeta, K^{-1}\eta\rangle_X = \chi(\vartheta).$$

Since K^{-1} is typically not continuous, ψ will not be continuous along the path $t \mapsto K\vartheta_t = \eta + tKb$ without further assumptions. Supposing however $\zeta \in \operatorname{ran} K^\star$, where $\operatorname{ran} K^\star$ denotes the range of the adjoint operator $K^\star$, continuity and linearity of ψ follows from

$$\psi(\eta) = \langle K^\star y, K^{-1}\eta\rangle_X = \langle y,\eta\rangle_Y \quad \text{for any } y \in (K^\star)^{-1}(\{\zeta\}).$$

In fact, we will see below that the condition $\zeta \in \operatorname{ran} K^\star$ is equivalent to the regularity of ψ. If $K^\star$ is not injective there are many solutions y of the equation $K^\star y = \zeta$. The unique solution with minimal norm is given by the Moore–Penrose pseudoinverse

$$(K^\star)^\dagger\zeta := (K^\star|_{(\ker K^\star)^\perp})^{-1} = \Pi_{(\ker K^\star)^\perp}(K^\star)^{-1}(\{\zeta\}) \quad \text{for} \quad \zeta \in \operatorname{ran} K^\star, \tag{3.3}$$

where $\Pi_{(\ker K^\star)^\perp}$ is the orthogonal projection onto the orthogonal complement $(\ker K^\star)^\perp$ of the kernel of $K^\star$, cf. Engl et al. (1996, Def. 2.2 and Prop. 2.3) for the definition and fundamental properties of the pseudoinverse.

Given the regularity of the parameter, a LAN version of the Hájek–Le Cam convolution theorem (see van der Vaart and Wellner, 1996, Thm. 3.11.2) yields that the variance of any regular estimator is bounded from below by

$$\|\Pi_{\overline{\operatorname{ran}}K}(K^\star)^{-1}(\{\zeta\})\|_Y^2 = \|(K^\star)^\dagger\zeta\|_Y^2 \quad \text{if} \quad \zeta \in \operatorname{ran} K^\star, \tag{3.4}$$

since the closure of the range of K satisfies $\overline{\operatorname{ran}}K = (\ker K^\star)^\perp$.

Remark 3.2. Suppose that the operator K is injective and compact and denote the domain of K by $\operatorname{dom} K$. Then K is adapted to the Hilbert scale $(\operatorname{dom}(K^\star K)^{-\alpha})_{\alpha\geqslant 0}$ generated by $(K^\star K)^{-1}$ and its degree of ill-posedness is $\alpha = 1/2$ (cf. Natterer, 1984).

According to Goldenshluger and Pereverzev (2000), the parameter $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X$ can be estimated with parametric rate ε if and only if $\vartheta \in \operatorname{ran}((K^\star K)^{1/2})$. Noting that $\operatorname{ran} K^\star = \operatorname{ran}((K^\star K)^{1/2})$ (Engl et al., 1996, Prop. 2.18), we recover the condition $\zeta \in \operatorname{ran} K^\star$. Since the existence of a regular estimator implies in particular that $\chi(\vartheta)$ can be estimated with rate ε, this condition is natural for stating a convolution theorem.

Remark 3.3. The information bound in (3.4) is sharp. Usually, regularization methods are necessary to construct estimators in ill-posed problems because $K^\dagger$ is unbounded and the observation $y_{\varepsilon,\vartheta}$ may not be in its domain. Assuming $\zeta \in \operatorname{ran} K^\star$, we can however define the estimator $\widehat{\chi}(\vartheta) := \langle (K^\star)^\dagger \zeta, y_{\varepsilon,\vartheta} \rangle_Y$, which satisfies

$$\begin{aligned}\widehat{\chi}(\vartheta) - \chi(\zeta) &= \langle (K^\dagger)^\star \zeta, K\vartheta \rangle_Y - \langle \zeta, \vartheta \rangle_X + \varepsilon \dot{W}((K^\dagger)^\star \zeta) \\ &= \langle \zeta, (K^\dagger K - \mathrm{Id})\vartheta \rangle_X + \varepsilon \dot{W}((K^\dagger)^\star \zeta) \sim \mathcal{N}(0, \varepsilon^2 \|(K^\dagger)^\star \zeta\|_Y^2)\end{aligned}$$

where we used $K^\dagger K = \Pi_{(\ker K)^\perp} = \mathrm{Id}$ because K is assumed to be injective. Therefore, the estimator $\widehat{\chi}(\vartheta)$ is efficient.

If the operator K in model (3.1) is nonlinear, the situation is more involved and a naive approach may fail as the following example illustrates.

Example 3.4. For a given $\vartheta \in \Theta := \{f \in L^2(\mathbb{R}) | f \geqslant 0\} \subseteq X := L^2(\mathbb{R})$ consider the linear differential equation in f

$$f' = -f + \vartheta^2 \quad \text{with} \quad \lim_{t \to -\infty} f(t) = 0, \tag{3.5}$$

which has the explicit solution $f_\vartheta(t) := \int_{-\infty}^t e^{-(t-s)} \vartheta^2(s) \mathrm{d}s$. The inverse problem is to estimate a linear functional $\chi(\vartheta) = \langle \vartheta, \zeta \rangle_X, \zeta \in X$, given an observation of the solution $f_\vartheta \in Y := L^2(\mathbb{R})$ of the previous equation corrupted by white noise. Since (3.5) is equivalent to $\mathcal{F}[\vartheta^2] = \mathcal{F}[f + f'] = (1 - iu)\,\mathcal{F} f$, the operator $K\colon \Theta \to L^2(\mathbb{R})$, which maps ϑ to the solution f_ϑ, can be written as

$$K(\vartheta) = \mathcal{F}^{-1}\left[(1 - iu)^{-1}\,\mathcal{F}[\vartheta^2](u)\right].$$

Note that K is well defined because $\|(1 - i\bullet)^{-1}\,\mathcal{F}[\vartheta^2]\|_{L^2} \leqslant \|(1 - i\bullet)^{-1}\|_{L^2(\mathbb{R})} \|\vartheta\|_{L^2(\mathbb{R})}^2$. We immediately see that K is nonlinear and injective on Θ. Due to the derivative in (3.5), ϑ does not depend continuously on the data f_ϑ and thus the problem is ill-posed.

Following the strategy of the linear model, we introduce the direct parameter $\eta = K(\vartheta)$ and write $\psi(\eta) = \langle \zeta, K^{-1}(\eta) \rangle_X = \chi(\vartheta)$. Note that ψ is nonlinear in η. To study pathwise continuity of ψ, we consider the path $[0, 1) \ni t \mapsto \eta_t = \eta + th$ with direction $h = K(b), b \in \Theta$. Note that $\eta_t \in \operatorname{ran} K$ since

$$(\eta_t + \eta_t')^{1/2} = (\eta + \eta' + t(h + h'))^{1/2} = (\vartheta^2 + tb^2)^{1/2} \in \Theta.$$

For some intermediate point $\xi \in [0, t]$ the mean value theorem yields

$$\begin{aligned} & t^{-1}(\psi(\eta_t) - \psi(\eta)) = t^{-1} \langle K^{-1}(\eta_t) - K^{-1}(\eta), \zeta \rangle_X \\ =& \langle \tfrac{1}{2}(\eta + \eta')^{-1/2}(h + h'), \zeta \rangle_X - t \langle \tfrac{1}{4}(\eta_\xi + \eta_\xi')^{-3/2}(h + h')^2, \zeta \rangle_X. \end{aligned} \tag{3.6}$$

The first term is the linearization $\dot\psi_\eta(h) = \frac{1}{2}\langle(\eta+\eta')^{-1/2}(h+h'),\zeta\rangle_X = \frac{1}{2}\langle\vartheta^{-1}b^2,\zeta\rangle_X$, where we have to impose suitable conditions on ζ first to compensate the potentially non-integrable singularities of ϑ^{-1} and second to ensure continuity in h. But even if these conditions are satisfied, pathwise continuity of ψ may fail because the integrability problems in the remainder in (3.6) are more serious because b^4 is not integrable for every $b \in X$ and the singularities of $(\vartheta^2+\xi b^2)^{-3/2}$ are more restrictive.

What went wrong in Example 3.4? Regularity of the parameter ψ depends on two properties: (i) the choice of ζ and (ii) the directions and paths along which we want to show the regularity. In particular, the second point has to capture the inverse structure of the problem. The approach in the following section provides a solution to both problems. It gives a clear condition on ζ and it determines appropriate perturbations of the parameter, described by the tangent space.

3.1.2. Local linear weak approximation

Turning to a much more general model, the following definition will ensure that it behaves locally like the model (3.1) with a linear operator. Let Θ be a parameter set such that for any $\vartheta \in \Theta$ there is a *tangent set* $\dot\Theta_\vartheta$ that is a subset of a Hilbert space with scalar product $\langle\bullet,\bullet\rangle_\vartheta$ such that any element $b \in \dot\Theta_\vartheta$ is associated to a path $[0,\tau) \ni t \mapsto \vartheta_t \in \Theta_\vartheta$ starting at ϑ and for some $\tau > 0$. For the sake of brevity we suppress the dependence of the path on b in the notation. In the following $Y_n \overset{P_n}{\Rightarrow} Y$ denotes weak convergence of the law of Y_n under the measure P_n to the law of Y for random variables $Y_1, Y_2, \dots$ and Y.

Definition 3.5. The sequence of statistical experiments $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ is called a *locally regular indirect model* at $\vartheta \in \Theta$ with respect to the tangent set $\dot\Theta_\vartheta$ if there are a Hilbert space $(H_\vartheta, \langle\bullet,\bullet\rangle_H)$ and a continuous linear operator

$$A_\vartheta \colon \operatorname{lin}\dot\Theta_\vartheta \to H_\vartheta$$

such that for some rate $r_n \downarrow 0$ and for every $b \in \dot\Theta_\vartheta$ with associated path $t \mapsto \vartheta_t$ there are random variables $(G_n(h))_{h\in\operatorname{ran}A_\vartheta}$ satisfying

$$\log\frac{\mathrm{d}P_{n,\vartheta_{r_n}}}{\mathrm{d}P_{n,\vartheta}} = G_n(A_\vartheta b) - \frac{1}{2}\|A_\vartheta b\|_H^2 \quad \text{and} \tag{3.7}$$

$$(G_n(h_1),\dots,G_n(h_k)) \overset{P_{n,\vartheta}}{\Longrightarrow} (G(h_1),\dots,G(h_k)) \text{ for all } k\in\mathbb{N}, h_1,\dots h_k \in \operatorname{ran}A_\vartheta,$$

for a centered Gaussian process $(G(h))_{h\in\operatorname{ran}A_\vartheta}$ with covariance structure given by $\mathbb{E}[G(h_1)G(h_2)] = \langle h_1,h_2\rangle_H$. The operator A_ϑ is called *generalized score operator*.

In the sequel we will use the notation

$$\mathcal{P}_n := \{P_{n,\vartheta} \,|\, \vartheta \in \Theta\}.$$

The statistical interpretation of this regularity becomes clear by comparing it to the likelihood ratio (3.2) in the linear white noise model. Condition (3.7) means that locally at ϑ the model $(P_{n,\vartheta_{r_n}})$ converges to a limit experiment which is a linear inverse problem

(3.1) in white noise with operator $K = A_\vartheta$ on the Hilbert space H_ϑ and with noise level $\varepsilon_n = r_n$. In other words, at ϑ the model converges weakly to the linear inverse problem in the sense of Le Cam (1972). Therefore, the classical white noise model (3.1) serves as a *locally linear weak approximation* of the general model $\mathcal{P}_n$. The difference to the classical theory is that the limit experiment is not a direct Gaussian shift experiment, but an indirect Gaussian shift, preserving the inverse structure of the problem. In that sense property (3.7) generalizes the classical local asymptotic normality, which corresponds to the identity operator $A_\vartheta = \mathrm{Id}$, to *local asymptotic indirect normality (LAIN)*.

The derived parameter $\chi : \Theta \to \mathbb{R}^d$, which is the aim of the statistical inference, should then be regular in the following sense.

Definition 3.6. The function $\chi : \Theta \mapsto \mathbb{R}^d, d \in \mathbb{N}$, is *pathwise differentiable* at $\vartheta \in \Theta$ with respect to the tangent set $\dot{\Theta}_\vartheta$ if there is a continuous linear operator $\dot{\chi}_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to \mathbb{R}^d$ such that for every $b \in \dot{\Theta}_\vartheta$ with associated path $[0, \tau) \ni t \mapsto \vartheta_t \in \Theta$ it holds

$$\frac{1}{t}(\chi(\vartheta_t) - \chi(\vartheta)) \to \dot{\chi}_\vartheta b \quad \text{as } t \downarrow 0.$$

By the Riesz representation theorem we can write $\dot{\chi}_\vartheta b = \langle \widetilde{\chi}_\vartheta, b \rangle_\vartheta$ for all $b \in \dot{\Theta}_\vartheta$ and some gradient $\widetilde{\chi}_\vartheta \in \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$. Recall that the sequence of parameter functions $\psi_n : \mathcal{P}_n \to \mathbb{R}^d$ is called regular at ϑ relative to $A_\vartheta \dot{\Theta}_\vartheta$ if for any $h \in A_\vartheta \dot{\Theta}_\vartheta$ and any submodel $t \mapsto P_{n,\vartheta_t}$ satisfying (3.7) with $h = A_\vartheta b$ for some $b \in \dot{\Theta}_\vartheta$, it holds

$$\frac{\psi_n(P_{n,\vartheta_{r_n}}) - \psi_n(P_{n,\vartheta})}{r_n} \to \dot{\psi}_\vartheta(h) \tag{3.8}$$

for some continuous linear map $\dot{\psi}_\vartheta : H \to \mathbb{R}^d$. Again the Riesz representation theorem determines a unique $\widetilde{\psi}_\vartheta \in \overline{\operatorname{ran}} A_\vartheta = \overline{\operatorname{lin}} A_\vartheta \dot{\Theta}_\vartheta$ such that $\dot{\psi}_\vartheta(h) = \langle \widetilde{\psi}_\vartheta, h \rangle_H$ for all $h \in \operatorname{ran} A_\vartheta$. $\widetilde{\psi}_\vartheta$ is called *efficient influence function* in the classical semiparametric theory. As the last ingredient we recall the following definition:

Definition 3.7. A sequence of estimators $T_n : \mathcal{X}_n \to \mathbb{R}^d$ is called *regular* at ϑ with respect to the rate r_n and relative to the directions $\dot{\Theta}_\vartheta$ if there is a limit distribution L on the Borel measurable space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that

$$\frac{1}{r_n}(T_n - \chi(\vartheta_{r_n})) \overset{P_{n,\vartheta_{r_n}}}{\Longrightarrow} L$$

for every $b \in \dot{\Theta}_\vartheta$ and any corresponding submodel $t \mapsto P_{n,\vartheta_t}$.

We recall the definition (3.3) of the Moore–Penrose pseudoinverse $K^\dagger$ of an operator K on its range and obtain the following convolution theorem.

Theorem 3.8. *Let $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ be a locally regular indirect model at $\vartheta \in \Theta$ and $\chi : \Theta \to \mathbb{R}^d$ be pathwise differentiable at ϑ with respect to $\dot{\Theta}_\vartheta$. Then the sequence $\psi_n : \mathcal{P}_n \to \mathbb{R}^d$ is regular at ϑ relative to $\dot{\Theta}_\vartheta$ if and only if each coordinate function of $\widetilde{\chi}_\vartheta = (\widetilde{\chi}_\vartheta^{(1)}, \dots, \widetilde{\chi}_\vartheta^{(d)})$ is contained in the range of the adjoint score operator $A_\vartheta^\star : H \to \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$.*

In this case $\widetilde{\psi}_\vartheta = (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta = ((A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(1)}, \dots, (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(d)})$ *is the efficient influence function and the limit distribution* L *of* $(T_n - \chi(\vartheta_{r_n}))/r_n$ *satisfies* $L = \mathcal{N}(0, \Sigma) * M$ *for some Borel probability distribution* M *and covariance matrix* $\Sigma \in \mathbb{R}^{d \times d}$ *given by*

$$\Sigma_{k,l} = \langle \widetilde{\psi}_\vartheta^{(k)}, \widetilde{\psi}_\vartheta^{(l)} \rangle_H = \langle (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(k)}, (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(l)} \rangle_H, \quad k, l \in \{1, \dots, d\}. \tag{3.9}$$

Proof. The characterization of regular parameter functions can be proved analogously to the i.i.d. setting studied by van der Vaart (1991, Thm. 3.1). Let us first consider $d = 1$. For any $b \in \dot{\Theta}_\vartheta$ there is a path $[0, \tau) \ni t \mapsto \vartheta_t$ in direction b generating a submodel $t \mapsto P_{n,\vartheta_t}$ with $h = A_\vartheta b$. If ψ_n is regular,

$$\dot{\psi}_\vartheta(A_\vartheta b) = \lim_{n\to\infty} \frac{\psi_n(P_{n,\vartheta_{r_n}}) - \psi_n(P_{n,\vartheta})}{r_n} = \lim_{n\to\infty} \frac{\chi(\vartheta_{r_n}) - \chi(\vartheta)}{r_n} = \langle \widetilde{\chi}_\vartheta, b \rangle_\vartheta.$$

Since the equality $\langle \widetilde{\psi}_\vartheta, A_\vartheta b \rangle_H = \dot{\psi}_\vartheta(A_\vartheta b) = \langle \widetilde{\chi}_\vartheta, b \rangle_\vartheta$ holds for all $b \in \dot{\Theta}_\vartheta$, it follows $A_\vartheta^\star \widetilde{\psi}_\vartheta = \widetilde{\chi}_\vartheta$. To conclude the converse direction, we can use the previous display as definition of $\dot{\psi}_\vartheta$ and have to verify that it is indeed linear and continuous. But this follows because by assumption there is some $\psi \in H$ such that $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$ and thus $\dot{\psi}_\vartheta(h) = \langle \psi, h \rangle_H$ for all $h \in H$. Because $\overline{\operatorname{ran}} A_\vartheta = (\ker A_\vartheta^\star)^\perp$, there is exactly one solution of $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$ in $\overline{\operatorname{ran}} A_\vartheta$ and this is given by $(A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta = \Pi_{(\ker A_\vartheta^\star)^\perp} (A_\vartheta^\star)^{-1}(\{\widetilde{\chi}_\vartheta\})$. For $d > 1$ it is sufficient to consider the coordinate functions separately.

To conclude the second part of the theorem, we consider $A_\vartheta \dot{\Theta}_\vartheta$ as local parameter set and identify the local parameter $\kappa_n(A_\vartheta b) := \psi_n(P_{n,\vartheta_{r_n}}) = \chi(\vartheta_{r_n})$ with $\kappa_n(0) := \psi_n(P_{n,\vartheta})$. Then κ_n is regular relative to $A_\vartheta \dot{\Theta}_\vartheta$ and we can apply the convolution theorem in van der Vaart and Wellner (1996, Thm. 3.11.2). Hereby, we have to note that it is sufficient if the local parameter set $A_\vartheta \dot{\Theta}_\vartheta$ is only a subset of a Hilbert space, and thus (3.7) may not hold for all linear combinations of elements in $\dot{\Theta}_\vartheta$, as long as the weak convergence $G_n(h) \Rightarrow G(h)$ under $P_{n,\vartheta}$ holds true for all $h \in \operatorname{lin} A_\vartheta \dot{\Theta}_\vartheta$. □

The theorem immediately implies that the asymptotic covariance of every regular estimator of $\chi(\vartheta)$ is bounded from below by (3.9) in the order of nonnegative definite matrices. If $\widetilde{\chi}_\vartheta$ is contained in the smaller range of the *information operator* $A_\vartheta^\star A_\vartheta$, then the efficient influence function can be obtained by

$$\widetilde{\psi}_\vartheta = (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta = A_\vartheta (A_\vartheta^\star A_\vartheta)^\dagger \widetilde{\chi}_\vartheta,$$

owing to $\ker(A_\vartheta^\star A_\vartheta) = \ker A_\vartheta$. Therefore, in this case the hardest parametric subproblem is given by the direction $(A_\vartheta^\star A_\vartheta)^\dagger \widetilde{\chi}_\vartheta \in \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$. In the finite dimensional linear model this lower bound coincides with the minimal variance of the Gauß–Markov theorem. Let us illustrate Theorem 3.8 in several examples.

Example 3.9 (Indirect regression model)**.** With $X = Y = L^2(\mathbb{R})$ and a bounded, linear and injective operator $K : X \to Y$ consider the indirect regression model with deterministic design

$$Y_i = (Kf)\Big(\frac{i}{n}\Big) + \varepsilon_i, \quad i = 1, \dots, n, \quad \text{with unkown } f \in X$$

and with i.i.d. errors $\varepsilon_1, \dots, \varepsilon_n \sim \mu$ for some law μ. Therefore, $(Y_1, \dots, Y_n) \sim P_{n,f} = \Pi_{i=1}^n \mu(\bullet - (Kf)(i/n))$. Khoujmane et al. (2007) have proved a convolution theorem for estimating $\chi(f) = \langle \zeta, f\rangle_{L^2(\mathbb{R})}, \zeta \in X$ with $\|\zeta\|_{L^2(\mathbb{R})} = 1$, where f and the error distribution μ are unknown. Let us focus on the submodel with standard normal errors $\varepsilon_i \sim \mathcal{N}(0,1)$. Under smoothness conditions it is stated in this submodel (Khoujmane et al., 2007, Thm. 2 with $\delta = 0$) that the asymptotic distribution of any regular estimator is given by the convolution $\mathcal{N}(0, \|K\zeta\|^{-2}_{L^2(\mathbb{R})}) * M$ for some Borel probability measure M.

To apply Theorem 3.8, we have to verify that this model is a regular indirect model. Choosing the tangent space $\dot{\Theta}_f = X$ with linear paths $f_t = f + tb$ in directions $b \in \dot{\Theta}_f$, we calculate

$$\begin{aligned}
&\log \frac{\mathrm{d}P_{n,f_{1/\sqrt{n}}}}{\mathrm{d}P_{n,f}}(Y_1,\dots,Y_n) \\
&= -\frac{1}{2n}\sum_{i=1}^n (Kb)^2\Big(\frac{i}{n}\Big) + \frac{1}{n^{1/2}}\sum_{i=1}^n (Kb)\Big(\frac{i}{n}\Big)\Big(Y_i - (Kf)\Big(\frac{i}{n}\Big)\Big) \\
&\xRightarrow{P_{n,f}} \mathcal{N}\Big(-\frac{1}{2}\|Kb\|^2_{L^2(\mathbb{R})}, \|Kb\|^2_{L^2(\mathbb{R})}\Big).
\end{aligned}$$

Therefore, the generalized score operator is given by $A_f = K$. Assuming for simplicity that $\zeta \in \operatorname{ran}(K^\star K)$, the asymptotic distribution of any regular estimator is given by a convolution $\mathcal{N}(0, \|K(K^\star K)^\dagger\zeta\|^2_{L^2(\mathbb{R})}) * M$. Since the Cauchy–Schwarz inequality yields $\|K\zeta\|_{L^2(\mathbb{R})}\|K(K^\star K)^\dagger\zeta\|_{L^2(\mathbb{R})} \geqslant \|\zeta\|^2_{L^2(\mathbb{R})} = 1$, the bound by Khoujmane et al. (2007) achieves our information bound if and only if $K^\star K = \lambda\,\mathrm{Id}$ for some $\lambda > 0$. Therefore, their information bound may not be optimal. The reason is that f has been perturbed in direction ζ instead of the least favorable direction $(K^\star K)^\dagger\zeta$.

Example 3.10 (Nonlinear white noise model)**.** Suppose we observe $y_{n,\vartheta} = K(\vartheta) + \varepsilon_n \dot{W}$ with $\varepsilon_n \to 0$ as $n \to \infty$ on the Hilbert space Y for some $\vartheta \in X$ and for a not necessarily linear operator $K : X \supseteq \Theta \to Y$ which is Gâteaux differentiable at the inner point $\vartheta \in \Theta$. That is there is a continuous linear operator $\dot{K}_\vartheta : X \to Y$ with

$$\lim_{t\to 0}\frac{1}{t}\big(K(\vartheta + tb) - K(\vartheta)\big) = \dot{K}_\vartheta b \quad \text{for all} \quad b \in X.$$

By the Hilbert space structure, the tangent space can be chosen as $\dot{\Theta}_\vartheta = X$ by considering the path $[0,1) \ni t \mapsto \vartheta_t := \vartheta + tb$ for $b \in \dot{\Theta}_\vartheta$. Lemma 3.1 yields for any $b \in X$ with associated path $t \mapsto \vartheta_t$

$$\log\frac{\mathrm{d}P_{n,\vartheta_{\varepsilon_n}}}{\mathrm{d}P_{n,\vartheta}}(y_{n,\vartheta}) = \dot{W}\Big(\frac{K(\vartheta+\varepsilon_n b) - K(\vartheta)}{\varepsilon_n}\Big) - \frac{1}{2\varepsilon_n^2}\|K(\vartheta+\varepsilon_n b) - K(\vartheta)\|_Y^2.$$

Therefore, the LAIN property (3.7) is satisfied with generalized score operator chosen as the Gâteaux derivative $A_\vartheta = \dot{K}_\vartheta$ at ϑ and

$$G_n(A_\vartheta b) = \dot{W}\Big(\frac{K(\vartheta+\varepsilon_n b) - K(\vartheta)}{\varepsilon_n}\Big) - \frac{1}{2}\Big(\Big\|\frac{K(\vartheta+\varepsilon_n b) - K(\vartheta)}{\varepsilon_n}\Big\|_Y^2 - \|\dot{K}_\vartheta b\|_Y^2\Big)$$

where $G_n(A_\vartheta b) \Rightarrow \mathcal{N}(0, \|\dot{K}_\vartheta b\|_Y^2)$, since the variance of the first term of G_n converges to $\|\dot{K}_\vartheta b\|_Y^2$ and the second term converges deterministically to zero by the Gâteaux differentiability. The convergence of the finite dimensional distributions follows likewise.

Along the path $t \mapsto \vartheta_t$ the linear functional $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X$ possesses the derivative

$$\lim_{t \to 0} \frac{1}{t} \big(\chi(\vartheta_t) - \chi(\vartheta) \big) = \langle \zeta, b \rangle_X.$$

Hence, $\dot{\chi}_\vartheta b = \langle \widetilde{\chi}_\vartheta, b \rangle_X$ for the gradient $\widetilde{\chi}_\vartheta = \zeta$ and all $b \in \dot{\Theta}_\vartheta$. Applying Theorem 3.8 shows in particular that the asymptotic variance of every regular sequence of estimators T_n (with respect to the rate ε_n) of the functional $\chi(\vartheta)$ is bounded from below by

$$\|(\dot{K}_\vartheta^\star)^\dagger \zeta\|_Y^2 \quad \text{whenever} \quad \zeta \in \operatorname{ran} \dot{K}_\vartheta^\star.$$

If K is a linear bounded operator the score operator is $A_\vartheta = \dot{K}_\vartheta = K$ and thus the statement of Theorem 3.8 coincides with the previous result (3.4).

In Remark 3.3 we saw that this information bound can be achieved if K is linear. For nonlinear operators an upper bound is still open.

Example 3.11 (Example 3.4 continued)**.** Let us come back to the ill-posed inverse problem in Example 3.4 related to the differential equation (3.5). The corresponding nonlinear operator $K \colon X \to Y$ with $\Theta = \{f \in L^2(\mathbb{R}) | f \geqslant 0\} \subseteq X$ and $X = Y = L^2(\mathbb{R})$ was given by $K(\vartheta) = \mathcal{F}^{-1}[(1 - i\bullet)^{-1} \mathcal{F}[\vartheta^2]]$. For $\vartheta \in \Theta$ and any $b \in \dot{\Theta}_\vartheta := \Theta$ the functional $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X, \zeta \in L^2(\mathbb{R})$, is pathwise differentiable along the path $[0,1) \mapsto \vartheta_t = \vartheta + tb$ with gradient $\widetilde{\chi}_\vartheta = \zeta$. K is pathwise differentiable with respect to the tangent set $\dot{\Theta}_\vartheta$ at ϑ with derivative

$$\dot{K}_\vartheta b = \mathcal{F}^{-1} \left[(1 - iu)^{-1} \mathcal{F}[2b\vartheta](u) \right].$$

Since $\dot{K}$ is well defined on $\operatorname{lin} \dot{\Theta}_\vartheta = L^2(\mathbb{R})$, the generalized score operator $A_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to H := L^2(\mathbb{R})$ is given by $A_\vartheta b = \dot{K}_\vartheta b$ as in the previous example. The "directions" in which we perturb the direct parameter $K(\vartheta)$ are then given by $A_\vartheta \dot{\Theta}_\vartheta = \{K(\sqrt{\vartheta b}) | b \in X\}$. Applying Plancherel's identity twice, the adjoint of A_ϑ can be calculated via

$$\begin{aligned} \langle \dot{K}_\vartheta b, h \rangle &= \frac{1}{2\pi} \int (1 - iu)^{-1} \mathcal{F}[2b\vartheta](u) \overline{\mathcal{F} h}(u) \mathrm{d}u \\ &= \langle b, 2\vartheta \, \mathcal{F}^{-1} \left[(1 - iu)^{-1} \mathcal{F} h(-u) \right] \rangle_X, \end{aligned}$$

for $b, h \in L^2(\mathbb{R})$. Therefore, $A_\vartheta^\star h = 2\vartheta \, \mathcal{F}^{-1}[(1 - iu)^{-1} \mathcal{F} h(-u)]$ and regularity of the parameter function follows for any $\zeta \in \operatorname{ran} A_\vartheta^\star = \{\vartheta f | f \in H^1(\mathbb{R})\}$ with the Sobolev space $H^1(\mathbb{R})$, cf. (A.1).

As Example 3.11 indicates, the adjoint score operator $A_\vartheta^\star$ has usually no closed range. In these cases it is a difficult problem to determine the range of $A_\vartheta^\star$. As the following characterization shows, it is sufficient to know $A_\vartheta^\star$ on a dense subspace. This approximation argument will turn out to be very useful for more complex models.

Proposition 3.12. *Let $A_\vartheta^\star : H \to \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$ be injective and let $\mathcal{G}$ be a dense subspace in H. Then $\widetilde{\chi}_\vartheta \in \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$ is contained in* $\operatorname{ran} A_\vartheta^\star$ *if and only if the following is satisfied*

(i) there exists a sequence $\chi_n \in \operatorname{ran} A_\vartheta^\star|_{\mathcal{G}}$ such that $\chi_n \to \widetilde{\chi}_\vartheta$ as $n \to \infty$ and

(ii) $(A_\vartheta^\star)^{-1} \chi_n$ converges to some $\psi \in H$.

In this case $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$ *and thus* $\Pi_{\overline{\operatorname{ran}} A_\vartheta} \psi = \psi$ *is the efficient influence function.*

Proof. "if": Since $A_\vartheta^\star$ is a bounded operator, its graph

$$\{(g, A_\vartheta^\star g) : g \in H\} \subseteq H \times \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$$

is closed. Therefore, the inverse operator $(A_\vartheta^\star)^{-1}|_{\operatorname{ran} A_\vartheta^\star}$ is closed, too. Consequently, (i) and (ii) imply $\widetilde{\chi}_\vartheta \in \operatorname{dom}(A_\vartheta^\star)^{-1} = \operatorname{ran} A_\vartheta^\star$ with $(A_\vartheta^\star)^{-1} \widetilde{\chi}_\vartheta = \psi$.

"only if": Since $\widetilde{\chi}_\vartheta \in \operatorname{ran} A_\vartheta^\star$ there is some $\psi \in H$ such that $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$. Moreover, there is a sequence $(g_n) \subseteq \mathcal{G}$ with $g_n \to \psi$ because $\mathcal{G}$ is dense in H. The continuity of $A_\vartheta^\star$ yields then $\chi_n := A_\vartheta^\star g_n \to A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$. □

3.1.3. I.i.d. observations

When the observations are given by n independent and identically distributed random variables $Y_1, \dots, Y_n$, the model simplifies to the product space $(\mathcal{X}^n, \mathcal{A}^{\otimes n}, P_\vartheta^{\otimes n} : \vartheta \in \Theta)$ such that the probability measure is completely described by the family of marginal distributions $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$. We will rephrase the conditions of the previous section in terms of the marginal measure P_ϑ. This setting appears quite often in applications and, in particular, the deconvolution model and the Lévy model which we study in Sections 3.2 and 3.3 will be two examples. That is why, we will give some details for the i.i.d. case.

Recall that a *tangent set* $\dot{\mathcal{P}}_{P_\vartheta}$ at P_ϑ is a set of score functions g of submodels $[0, \tau) \ni t \mapsto P_{\vartheta_t}$ starting at P_ϑ and for some $\tau > 0$. In the present situation the derived parameter can be written as $\psi(P_\vartheta) = \chi(\vartheta)$, independent of n. The classical Hajék–Le Cam convolution theorem (cf. Bickel et al., 1998, Thm. 3.3.2) applies if ψ is differentiable at P_ϑ relative to $\dot{\mathcal{P}}_{P_\vartheta}$, that is, there exists a continuous linear map $\dot{\psi} : L^2(P_\vartheta) \to \mathbb{R}^k$ such that

$$\lim_{t \to 0} \frac{\psi(P_{\vartheta_t}) - \psi(P_\vartheta)}{t} = \dot{\psi} g \quad \text{for all } g \in \dot{\mathcal{P}}_{P_\vartheta}.$$

This differentiability corresponds to the general assumption (3.8). In the i.i.d. setting local asymptotic normality follows from Hellinger regularity of the submodel $t \mapsto P_{\vartheta_t}$. Therefore, we can reformulate the conditions in Definition 3.5 to the following:

Assumption 3.A. At $\vartheta \in \Theta$ let the parameter set give rise to a tangent set $\dot{\Theta}_\vartheta$. Furthermore, let there be a continuous linear operator

$$A_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to L_0^2(P_\vartheta) := \Big\{ g \in L^2(P_\vartheta) : \int g \mathrm{d}P_\vartheta = 0 \Big\}$$

such that for every $b \in \dot{\Theta}_\vartheta$ with associated path $t \mapsto \vartheta_t$

$$\int \Big(\frac{\mathrm{d}P_{\vartheta_t}^{1/2} - \mathrm{d}P_\vartheta^{1/2}}{t} - \frac{1}{2} A_\vartheta b \mathrm{d}P_\vartheta^{1/2} \Big)^2 \to 0 \quad \text{as } t \downarrow 0. \tag{3.10}$$

In (3.10) $\mathrm{d}P_{\vartheta_t}$ denotes the Radon–Nikodym μ-density of P_{ϑ_t} for some dominating measure μ and the integration is with respect to μ. Since the integral does not depend on μ, it is suppressed in the notation.

Lemma 3.13. *If the product model $(\mathcal{X}^n, \mathcal{A}^{\otimes n}, P_\vartheta^{\otimes n} : \vartheta \in \Theta)$ satisfies Assumption 3.A at $\vartheta \in \Theta$, then it is a locally regular indirect model at $\vartheta \in \Theta$ with respect to the tangent set $\dot{\Theta}_\vartheta$, with rate $r_n = n^{-1/2}$ and (generalized) score operator A_ϑ.*

Proof. The Hellinger regularity in Assumption 3.A implies local asymptotic normality since it yields, for instance, see Bickel et al. (1998, Prop. 2.1.2),

$$\sum_{j=1}^n \log \frac{\mathrm{d}P_{\vartheta_{1/\sqrt{n}}}}{\mathrm{d}P_\vartheta}(X_j) = \frac{1}{\sqrt{n}} \sum_{j=1}^n A_\vartheta b(X_j) - \frac{1}{2}\|A_\vartheta b\|_{L^2(P_\vartheta)} + R_n$$

for a remainder R_n that converges in $P_\vartheta^{\otimes n}$-probability to zero. Hence, the LAIN property (3.7) is statisfied with rate $1/\sqrt{n}$ and with the score operator A_ϑ mapping into the Hilbert space $H = L_0^2(P_\vartheta)$. The convergence of the finite dimensional distributions in Definition 3.5 follows from the Cramér–Wold device and the linearity of A_ϑ. □

Note that $L_0^2(P_\nu)$ is the orthogonal complement of $\operatorname{lin} 1$ and thus it is a closed subspace of the Hilbert space $L^2(P_\nu)$. The operator A_ϑ maps directions $b \in \dot{\Theta}_\vartheta$ into score functions at P_ϑ and thus it is called *score operator* which explains the name given in the general case. It generates the tangent set $\dot{\mathcal{P}}_{P_\vartheta} = A_\vartheta \dot{\Theta}_\vartheta$ of the model $\mathcal{P}$ at P_ϑ. Note that the range of A_ϑ is a subset of the maximal tangent set as the following example shows.

Example 3.14 (Maximal tangent set)**.** Let $\mathcal{P}$ be the model of all probability measures on some sample space. The maximal tangent set of the model $\mathcal{P}$ at some distribution P is given by $L_0^2(P)$. This can be seen as follows: Score functions are necessarily centered and square integrable. For any score $g \in L_0^2(P)$ a one-dimensional submodel is $t \to c(t)k(tg(x))\mathrm{d}P(x)$ with a twice continuously differentiable function $k\colon \mathbb{R} \to \mathbb{R}_+$ which satisfies $k(0) = k'(0) = 1$ and such that k'/k is bounded and with normalization constant $c(t) = \|k(tb)\|_{L^1(\nu)}^{-1}$, for instance, $k(y) = 2/(1+e^{-2y})$ (cf. van der Vaart, 1998, Ex. 25.16).

Theorem 3.8 yields then the following convolution theorem, which was already obtained by van der Vaart (1991).

Corollary 3.15. *Suppose the product model with marginal distributions $\mathcal{P} = \{P_\vartheta | \vartheta \in \Theta\}$ satisfies Assumption 3.A and let $\chi : \Theta \to \mathbb{R}^d$ be pathwise differentiable with respect to $\dot{\Theta}_\vartheta$. The map $\psi : \mathcal{P} \to \mathbb{R}^d$ is differentiable at P_ϑ relative to the tangent set $\dot{\mathcal{P}}_{P_\vartheta} = A_\vartheta \dot{\Theta}_\vartheta$ if and only if each coordinate function of $\widetilde{\chi}_\vartheta$ is contained in the range of the adjoint score operator $A_\vartheta^\star : L_0^2(P_\vartheta) \to \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$. In this case the efficient influence function is given by $\widetilde{\psi}_\vartheta = (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta$.*

In particular, for $\zeta \in \operatorname{ran} A_\vartheta^\star$ the asymptotic covariance matrix of every regular estimator is bounded from below by

$$\mathbb{E}_\vartheta \left[\widetilde{\psi}_{P_\vartheta} \widetilde{\psi}_{P_\vartheta}^\top\right] = \mathbb{E}_\vartheta \left[\big((A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta\big)\big((A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta\big)^\top\right].$$

If $\widetilde{\chi}_\vartheta \notin \operatorname{ran} A_\vartheta^\star$, van der Vaart (1991) shows that there exists no regular estimator of the functional $\chi(\vartheta)$.

In the i.i.d. case we find the following statistical interpretation of Proposition 3.12, adopting the Cramér–Rao point of view. Let $\mathcal{G}$ be a dense subset of $L_0^2(P_\vartheta)$ and let $\chi(\vartheta)$

be a one-dimensional derived parameter with gradient $\widetilde{\chi}_\vartheta$. Consider an approximating sequence $\chi_n \to \widetilde{\chi}_\vartheta$ satisfying $\mathcal{G} \ni g_n := (A_\vartheta^\star)^\dagger \chi_n \to \widetilde{\psi}_{P_\vartheta}$. Assuming $\mathcal{G} \subseteq \operatorname{ran} A_\vartheta$, we can define $b_n := A_\vartheta^\dagger g_n = I^\dagger \chi_n$ where $I := A_\vartheta^\star A_\vartheta$ is the information operator. The information bound can be read as a Cramér–Rao bound in the least favorable submodel

$$\begin{aligned}\mathbb{E}[\widetilde{\psi}_{P_\vartheta}^2] &= \sup_{g \in \operatorname{lin} \dot{\mathcal{P}}_{P_\vartheta}} \frac{\langle \widetilde{\psi}_{P_\vartheta}, g\rangle_{P_\vartheta}^2}{\langle g, g\rangle_{P_\vartheta}} = \sup_{b \in \operatorname{lin} \dot{\Theta}_\vartheta} \frac{\langle \widetilde{\chi}_\vartheta, b\rangle_\vartheta^2}{\langle A_\vartheta b, A_\vartheta b\rangle_{P_\vartheta}} \\ &\geqslant \frac{\big(\langle \chi_n, b_n\rangle_\vartheta - \langle \widetilde{\chi}_\vartheta - \chi_n, b_n\rangle_\vartheta\big)^2}{\langle A_\vartheta b_n, A_\vartheta b_n\rangle_{P_\vartheta}},\end{aligned} \tag{3.11}$$

where we plugged in the direction $b_n = I^\dagger \chi_n$. The term $\langle \chi_n, b_n\rangle_\vartheta^2 / \langle A_\vartheta b_n, A_\vartheta b_n\rangle_{P_\vartheta} = \langle g_n, g_n\rangle_{P_\vartheta}$ is the Cramér–Rao bound for the estimation problem of a functional, which approximates $\chi(\vartheta)$, with gradient χ_n. The approximation error $\langle \widetilde{\chi}_\vartheta - \chi_n, b_n\rangle_\vartheta$ should be understood as bias. Since b_n does not have to be bounded, $\chi_n \to \widetilde{\chi}_\vartheta$ is not sufficient to conclude that the bias vanishes. However, Proposition 3.12(ii) implies that this error converges to zero owing to the Cauchy–Schwarz inequality:

$$\begin{aligned}|\langle \widetilde{\chi}_\vartheta - \chi_n, b_n\rangle_\vartheta| &= |\langle (A_\vartheta^\star)^\dagger(\widetilde{\chi}_\vartheta - \chi_n), A_\vartheta b_n\rangle_{P_\vartheta}| \\ &\leqslant \|\widetilde{\psi}_{P_\vartheta} - g_n\|_{P_\vartheta} \|g_n\|_{P_\vartheta} \to 0.\end{aligned}$$

Hence, the Cramér–Rao bound (3.11) converges to the information bound $\langle \widetilde{\psi}_{P_\vartheta}, \widetilde{\psi}_{P_\vartheta}\rangle_{P_\vartheta}$. A similar perspective was taken by Söhl and Trabs (2012b, Lem. 3).

3.1.4. Extension to Banach space valued functions

So far we considered $\mathbb{R}^d$-valued derived parameters χ. The aim of this section is to generalize Theorem 3.8 to functions $\chi : \Theta \to \mathbb{B}$ for a Banach space $(\mathbb{B}, \|\bullet\|)$. As pointed out by van der Vaart (1988) for the estimation of parameters in infinite dimensional spaces efficiency means essentially efficiency for the marginals plus tightness of the limit law of the sequence of estimator.

Let $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ be a locally regular indirect model at $\vartheta \in \Theta$ with respect to the tangent set $\dot{\Theta}_\vartheta$ and with generalized score operator $A_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to H_\vartheta$ for some Hilbert space $(H_\vartheta, \langle \bullet, \bullet\rangle_H)$. First, we have to generalize the notion of regularity to Banach space valued functions. The derived parameter $\chi : \Theta \to \mathbb{B}$ is *pathwise differentiable* at $\vartheta \in \Theta$ with respect to the tangent set $\dot{\Theta}_\vartheta$ if for all $b \in \dot{\Theta}_\vartheta$ with associated path $[0, \tau) \ni t \mapsto \vartheta_t$

$$t^{-1}\big(\chi(\vartheta_t) - \chi(\vartheta)\big) \to \dot{\chi}_\vartheta b$$

holds true for some continuous linear map $\dot{\chi}_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to \mathbb{B}$. The gradient of $\dot{\chi}_\vartheta$ is then defined as by van der Vaart (1991) using the dual space $\mathbb{B}^\star$, which is the space of all continuous linear functions $b^\star : \mathbb{B} \to \mathbb{R}$. The composition $b^\star \circ \dot{\chi}_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to \mathbb{R}$ is linear and continuous and thus it can be represented by some $\widetilde{\chi}_{\vartheta, b^\star} \in \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$:

$$(b^\star \circ \dot{\chi}_\vartheta) b = \langle \widetilde{\chi}_{\vartheta, b^\star}, b\rangle_\vartheta \quad \text{for all } b \in \operatorname{lin} \dot{\Theta}_\vartheta.$$

Similarly, the parameter $\psi_n(P_{n,\vartheta}) = \chi(\vartheta)$ is *regular* if (3.8) holds for some continuous linear map $\dot{\psi}_\vartheta : H \to \mathbb{B}$. The efficient influence functions $\widetilde{\psi}_{\vartheta, b^\star} \in \overline{\operatorname{ran}} A_\vartheta$ are defined by

$(b^\star \circ \dot{\psi}_\vartheta)h = \langle \widetilde{\psi}_{\vartheta,b^\star}, h\rangle_H$ for all $h \in \operatorname{ran} A_\vartheta$. The sequence of estimators $T_n : \mathcal{X}_n \to \mathbb{B}$ is called *regular* at $\vartheta \in \Theta$ with respect to the rate r_n if there is a fixed tight Borel probability measure L on $\mathbb{B}$ such that for all $b \in \dot{\Theta}_\vartheta$ with corresponding submodel $t \mapsto P_{n,\vartheta_t}$

$$\frac{1}{r_n}(T_n - \chi(\vartheta_{r_n})) \overset{P_{n,\vartheta_{r_n}}}{\Longrightarrow} L,$$

where weak convergence is defined in terms of outer probability to avoid measurability problems, that is,

$$\mathbb{E}^*_{P_{n,\vartheta_{r_n}}}\left[f\left(r_n^{-1}(T_n - \chi(\vartheta_{r_n}))\right]\right. \to \int f \mathrm{d}L$$

for all bounded, continuous function $f : \mathbb{B} \to \mathbb{R}$ (cf. van der Vaart and Wellner, 1996, Def. 1.3.3). Now we can state the following generalization of Theorem 3.8.

Theorem 3.16. *Let $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ be a locally regular indirect model at $\vartheta \in \Theta$ with respect to $\dot{\Theta}_\vartheta$ and let $\chi : \Theta \to \mathbb{B}$ be pathwise differentiable at ϑ with respect to $\dot{\Theta}_\vartheta$. Then the sequence $\psi_n : \mathcal{P}_n \to \mathbb{B}$ is regular at ϑ relative to $\dot{\Theta}_\vartheta$ if and only if*

$$\widetilde{\chi}_{\vartheta,b^\star} \in \operatorname{ran} A_\vartheta^\star \quad \textit{for all} \quad b^\star \in \mathbb{B}^\star. \tag{3.12}$$

In this case the efficient influence functions are given by the Moore–Penrose pseudoinverse $\widetilde{\psi}_{\vartheta,b^\star} = (A_\vartheta^\star)^\dagger \widetilde{\chi}_{\vartheta,b^\star}$. For any regular sequence of estimators T_n the limit distribution L of $r_n^{-1}(T_n - \chi(\vartheta_{r_n}))$ is given by the law of a sum $N + W$ of independent, tight, Borel measurable random elements in $\mathbb{B}$ such that

$$b^\star N \sim \mathcal{N}(0, \|\widetilde{\psi}_{\vartheta,b^\star}\|_H^2) = \mathcal{N}(0, \|(A_\vartheta^\star)^\dagger \widetilde{\chi}_{\vartheta,b^\star}\|_H^2).$$

The proof of this Theorem is analogous to Theorem 3.8 with an additional application of Lemma A.2 by van der Vaart (1991). We omit the details.

Remark 3.17. The type of regularity which we used for the parameters $\psi_n(P_{n,\vartheta})$ to apply the convolution theorem is quite strong because the derivative $\dot{\psi}_\vartheta$ has to be continuous with respect to the norm topology of $\mathbb{B}$. Necessarily, the range condition (3.12) has to hold for all $b^\star \in \mathbb{B}^\star$ which may fail if the dual space is large. This problem can be solved by using a weaker topology on $\mathbb{B}$ which is generated by a subspace $B' \subseteq \mathbb{B}^\star$ as shown by van der Vaart (1988, Sect. 3).

To show tightness of the limit distribution may be a difficult problem for inverse problems. In the i.i.d. setting the theory of smoothed empirical processes by Giné and Nickl (2008) turns out to be useful as the articles by Nickl and Reiß (2012) and Nickl et al. (2013) show.

3.2. Deconvolution

Based on the previous results, we study semiparametric efficiency in the classical nonparametric deconvolution setup. Recall that we observe an i.i.d. sample

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \ldots, n, \tag{3.13}$$

where X_j and ε_j are independent and have distributions ν and μ, respectively. Let us assume in this section that μ is known and thus the model is

$$\mathcal{P} = \{P_\nu = \nu * \mu | \nu \in \Theta\} \quad \text{with} \quad \Theta := \{\nu \text{ probability measure on } (\mathbb{R}, \mathscr{B}(\mathbb{R}))\}.$$

The statistical problem is to estimate the linear functional $\psi(P_\nu) = \chi(\nu) := \int \zeta \mathrm{d}\nu$ with $\zeta \in L^2(\nu)$. One of the most interesting examples is the estimation of the distribution function of X_j, corresponding to $\zeta = \mathbb{1}_{(-\infty,t]}$ for $t \in \mathbb{R}$. Later in Chapter 4 we will in detail study minimax rates and adaptive estimation of this particular functional and the related quantiles.

3.2.1. Lower bound

we aim for a convolution theorem. In a general linear indirect density estimation setting, a convolution theorem was already proved by van Rooij et al. (1999), who use the spectral decomposition of the operator. Their approach applies however only for a restricted class of functionals, depending on the polar decomposition, and they need an abstract condition on the density of ν which is difficult to verify. It implicitly assumes an appropriate decay behavior on this density. Their application to the deconvolution setting is restricted to a specific example. Studying deconvolution in more detail, Söhl and Trabs (2012b) have shown an information bound, assuming a polynomial decay behavior of a sufficiently regular Lebesgue density of ν and a bit more than second moments. They described the class of admissible functionals analytically, including the estimation of the distribution function of ν. Here, we are able to relax the conditions on ν considerably.

For any $\nu \in \Theta$ we choose the tangent space

$$\dot{\Theta}_\nu := L^2_0(\nu) = \overline{\mathrm{lin}} \dot{\Theta}_\vartheta. \tag{3.14}$$

According to Example 3.14, $\dot{\Theta}_\nu$ coincides with the maximal tangent set for direct observations. For any direction $b \in \dot{\Theta}_\nu$ and some sufficiently small $\tau > 0$ the path $[0,\tau) \ni t \to \nu_t$ where $\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu} = k(tb)/\int k(tb)\mathrm{d}\nu$ with $k : \mathbb{R} \to \mathbb{R}_+$ as in Example 3.14 is a submodel of Θ with $b = \frac{\partial}{\partial t}|_{t=0} \log(\mathrm{d}\nu_t)$. Using $|k(tb)| \leqslant t|b| \in L^2(\nu)$ and dominated convergence, the pathwise derivative of χ along $t \mapsto \nu_t$ at $t = 0$ is given by

$$\begin{aligned}\lim_{t\to 0} t^{-1}(\chi(\nu_t) - \chi(\nu)) &= \lim_{t\to 0} \int \zeta(x) t^{-1}\Big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) - 1\Big)\mathrm{d}\nu(x)\\ &= \int \zeta(x)b(x)\nu(\mathrm{d}x) = \langle \zeta, b\rangle_\nu =: \dot{\chi}_\nu b.\end{aligned}$$

Hence, the derivative can be represented by $\dot{\chi}_\nu b = \langle \widetilde{\chi}_\nu, b\rangle_\nu$ for $\widetilde{\chi}_\nu = \zeta - \int \zeta \mathrm{d}\nu \in \dot{\Theta}_\nu$. The path $t \mapsto \nu_t$ induces a regular submodel $t \mapsto P_{\nu_t} = \nu_t * \mu$ which is shown by the following lemma.

Lemma 3.18. *For any nonzero $b \in \dot{\Theta}_\nu = L^2_0(\nu)$ the submodel $[0,\tau) \ni t \mapsto P_{\nu_t} = \nu_t * \mu$, for $\tau > 0$ sufficiently small, is Hellinger differentiable, that is* (3.10) *holds with continuous score operator*

$$A_\nu : \quad \dot{\Theta}_\nu \to L^2_0(P_\nu), \quad b \mapsto \mathbb{E}\left[b(X)|X+\varepsilon\right] = \frac{\mathrm{d}\big((b\nu) * \mu\big)}{\mathrm{d}P_\nu}, \tag{3.15}$$

where the expectation is taken with respect to the product measure measure $P^{(X,\varepsilon)} = \nu \otimes \mu$.

Proof. First, note that the (signed) measure $(f\nu) * \mu$ is absolutely continuous with respect to $\nu * \mu$ for any $f \in L^1(\nu)$, written as $(f\nu) * \mu \ll \nu * \mu$. In particular, the Radon–Nikodym density in (3.15) is well defined and $P_{\nu_t} \ll P_\nu$ for all $t > 0$. Let us denote $p_t(y) := \frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu}(y)$, $n_t(x) := \frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) = k(tb(x))/\int k(tb)\mathrm{d}\nu$ and write $\mathbb{E}_t[\bullet]$ for the expectation under $P_t^{(X,\varepsilon)} = \nu_t \otimes \mu$. Let $\mathbb{R} \times \mathscr{B}(\mathbb{R}) \ni (y, A) \mapsto \kappa_{X,X+\varepsilon}(y, A)$ be the regular conditional probability of $P^{(X,\varepsilon)}(X \in \bullet | X + \varepsilon)$ that is

$$\kappa_{X,X+\varepsilon}(y, A) = P^{(X,\varepsilon)}(X \in A | X + \varepsilon = y)$$

for P_ν-a.e. $y \in \mathbb{R}$ and all $A \in \mathscr{B}(\mathbb{R})$. We claim

$$p_t(Y) = \mathbb{E}_0\left[n_t(X) | X + \varepsilon = Y\right] = \int n_t(x)\kappa_{X,X+\varepsilon}(Y, \mathrm{d}x) \quad P_\nu - a.s. \tag{3.16}$$

To verify (3.16), we note for any Borel set $A \in \mathscr{B}(\mathbb{R})$

$$\begin{aligned}\mathbb{E}_0[\mathbb{1}_A(Y)p_t(Y)] = \mathbb{E}_t[\mathbb{1}_A(Y)] &= \mathbb{E}_0[\mathbb{1}_A(X+\varepsilon)n_t(X)] \\ &= \mathbb{E}_0\left[\mathbb{1}_A(Y)\,\mathbb{E}_0[n_t(X)|X+\varepsilon = Y]\right]\end{aligned} \tag{3.17}$$

which shows the first equality in (3.16). The second one follows from the choice of $\kappa_{X,X+\varepsilon}$.

We will show regularity of the submodel $(-\tau, \tau) \ni t \mapsto P_{\nu_t} = \nu_t * \mu$ for a sufficiently small $\tau > 0$ by applying Proposition 2.1.1 by Bickel et al. (1998). Using the properties of k,

$$\dot{n}_t(x) := \frac{\partial}{\partial t} n_t(x) = \frac{b(x)k'(tb(x))\int k(tb)\mathrm{d}\nu - k(tb(x))\int bk'(tb)\mathrm{d}\nu}{(\int k(tb)\mathrm{d}\nu)^2}$$

can be bounded uniformly in $t \in (-\tau, \tau)$ by $c_b(b(x)+1)$ for a constant $c_b > 0$, depending on b, and it is continuous in t on $(-\tau, \tau)$ for some sufficiently small $\tau > 0$. Since $b \in L^2(\nu) \subseteq L^2(\nu \otimes \mu)$, dominated convergence and (3.16) yield that $p_t(y)$ is continuously differentiable in $t \in (-\tau, \tau)$ for P_ν-almost all $y \in \mathbb{R}$ with derivative

$$\dot{p}_t(y) = \frac{\partial}{\partial t} p_t(y) = \int \dot{n}_t(x)\kappa_{X,X+\varepsilon}(y, \mathrm{d}x) = \mathbb{E}_0[\dot{n}_t(X)|X+\varepsilon = y].$$

By Jensen's inequality we see that

$$\begin{aligned}\|\dot{p}_t\|^2_{L^2(P_t)} &= \mathbb{E}_0\left[p_t(Y)\big|\,\mathbb{E}_0[\dot{n}_t(X)|X+\varepsilon = Y]\big|^2\right], \\ &\leqslant \mathbb{E}_0\left[p_t(Y)|\dot{n}_t(X)|^2\right].\end{aligned} \tag{3.18}$$

Since $n_t(x)$ can be bounded uniformly in $t \in (-\tau, \tau)$ and $x \in \mathbb{R}$, the density $p_t(Y)$ is P_ν-a.s. bounded by some constant $C > 0$ owing to (3.16). Therefore, we conclude from the previous estimate together with the bound $|\dot{n}_t| \leqslant c_b(b+1)$ and $b \in L^2(\nu)$ that

$$\|\dot{p}_t\|^2_{L^2(P_t)} \leqslant C\,\mathbb{E}_0[|\dot{n}_t(X)|^2] \leqslant Cc_b^2\,\mathbb{E}_0[(b(X)+1)^2] < \infty.$$

In particular, the Fisher information $I_t := \mathbb{E}_t[\dot{p}_t(Y)^2]$ is finite. Using (3.18), we infer that I_t is continuous. Noting that

$$\frac{\partial}{\partial t}\Big|_{t=0}\int k(tb)\mathrm{d}\nu = \int b\mathrm{d}\nu = 0 \quad \text{and thus} \quad \dot{n}_0(y) = \frac{\partial}{\partial t}\Big|_{t=0} n_t(x) = b(x),$$

we have $I_0 = \mathbb{E}_0[b(X)^2]$ and I_t is therefore nonzero for $b \neq 0$ and t small enough. In combination with the continuous differentiability of p_t, Proposition 2.1.1 by Bickel et al. (1998) yields the Hellinger differentiability (3.10) at $t = 0$ with derivative $\frac{1}{2}\dot{p}_0$. We obtain the score operator

$$A_\nu b := \dot{p}_0 = \mathbb{E}_0[b(X)|X + \varepsilon].$$

To see that $A_\nu : \dot{\Theta}_\nu \to L_0^2(P_\nu)$ is well defined and continuous, we again use Jensen's inequality which yields

$$\mathbb{E}_0[A_\nu b] = \mathbb{E}_0[b(X)] = 0 \quad \text{and} \quad \mathbb{E}_0[|A_\nu b|^2] \leqslant \mathbb{E}_0[|b(X)|^2] = \|b\|^2_{L^2(\nu)}.$$

Finally, a similar calculation as (3.17) shows $A_\nu b = \mathbb{E}_0[b(X)|X+\varepsilon] = \frac{\mathrm{d}((b\nu)*\mu)}{\mathrm{d}P_\nu}$. $\square$

Lemma 3.18 shows that Assumption 3.A is satisfied. In order to apply Corollary 3.15, we have to determine the adjoint of the score operator. For any $g \in L_0^2(P_\nu)$ and any $b \in \dot{\Theta}_\vartheta \subseteq L^2(\nu)$ the map $\mathbb{R}^2 \ni (x, y) \mapsto g(x + y)b(x)$ is $\nu \otimes \mu$-integrable due to the Cauchy–Schwarz inequality. Hence, $(b\nu) * \mu \ll P_\nu$ and Fubini's theorem yield

$$\begin{aligned}\langle A_\nu b, g\rangle_{P_\nu} &= \int (A_\nu b) g \mathrm{d}P_\nu = \int g\mathrm{d}((b\nu) * \mu) \\ &= \int\int g(x+y)b(x)\nu(\mathrm{d}x)\mu(\mathrm{d}y) = \langle \mu(-\bullet) * g, b\rangle_\nu.\end{aligned}$$

Noting that Jensen's inequality shows

$$\|\mu(-\bullet) * g\|^2_{L^2(\nu)} = \int\Big(\int g(x+y)\mu(\mathrm{d}y)\Big)^2\nu(\mathrm{d}x) \leqslant \int g^2\mathrm{d}(\nu * \mu) = \|g\|^2_{L^2(P_\nu)}$$

and that $\int(\mu(-\bullet) * g)\mathrm{d}\nu = \int g\mathrm{d}(\nu * \mu) = 0$, the adjoint score operator equals

$$A_\nu^\star : \quad L_0^2(P_\nu) \to \dot{\Theta}_\nu, \quad g \mapsto \mu(-\bullet) * g. \tag{3.19}$$

Under weak conditions on the measures ν and μ we conclude the following convolution theorem.

Theorem 3.19. *In the deconvolution model* (3.13) *suppose that* $\varphi_\varepsilon(u) := \mathbb{E}[e^{iu\varepsilon_1}] = \mathcal{F}\mu(u) \neq 0$ *for all* $u \in \mathbb{R}$ *and that* ν *admits a Lebesgue density. Then* $A_\nu^\star$ *as given in* (3.19) *is injective. Let moreover* $\zeta^{(1)}, \ldots, \zeta^{(d)} \in L^2(\nu)$ *satisfy* $\zeta^{(j)} \in \operatorname{ran} A_\nu^\star$ *for* $j = 1, \ldots, d$ *and* $d \in \mathbb{N}$. *Then the limit distribution of any regular estimator of the parameter* $(\int \zeta^{(1)}\mathrm{d}\nu, \ldots, \int \zeta^{(d)}\mathrm{d}\nu)$ *equals* $\mathcal{N}(0, \Sigma) * M$ *for some Borel probability measure* M *and with covariance matrix* $\Sigma \in \mathbb{R}^{d\times d}$ *given by*

$$\Sigma_{j,k} = \int \Big((A_\nu^\star)^{-1}\zeta^{(j)}\Big)\Big((A_\nu^\star)^{-1}\zeta^{(k)}\Big)\mathrm{d}P_\nu - \Big(\int \zeta^{(j)}\mathrm{d}\nu\Big)\Big(\int\zeta^{(k)}\mathrm{d}\nu\Big) \tag{3.20}$$

for $j, k = 1, \ldots, d$.

Proof. Let us first show that on the assumptions the adjoint operator $A_\nu^\star$ is injective. Since ν admits a Lebesgue density, the equivalence classes with respect to the Lebesgue measure embed into the equivalence classes with respect to ν and with respect to P_ν. Hence, we can consider the subset $\mathcal{G} := L^2(\mathbb{R}) \cap L_0^2(P_\nu)$, which is dense in $L_0^2(P_\nu)$ (cf. Section 3.4.3 (ii)). Since the kernel of the continuous operator $A_\nu^\star$ is closed, it is sufficient to show that the restricted operator $A_\nu^\star|_\mathcal{G}$ is injective. For any $g \in \mathcal{G}$ it holds $0 = A_\nu^\star g = \mu(-\bullet) * g$ if and only if $0 = \mathcal{F}[\mu(-\bullet) * g] = \varphi_\varepsilon(-\bullet)\,\mathcal{F}\,g$, which is equivalent to $\mathcal{F}\,g = 0$ since $|\varphi_\varepsilon| > 0$ by assumption. Hence, the kernel of $A_\nu^\star$ equals $\{0\}$.

To infer the information bound, recall that the gradient of the linear functional $\int \zeta^{(j)}\mathrm{d}\nu$ is $\widetilde{\chi}_\nu^{(j)} = \zeta^{(j)} - \int \zeta^{(j)}\mathrm{d}\nu$. Furthermore, note that $\zeta^{(j)} \in \operatorname{ran} A_\nu^\star$ implies $\widetilde{\chi}_\nu^{(j)} \in \operatorname{ran} A_\nu^\star$ since $A_\nu^\star a = \mu(-\bullet) * a = \int a\mathrm{d}\mu = a$ for any real number $a \in \mathbb{R}$. Therefore, Lemma 3.18, Corollary 3.15 and the injectivity of $A_\nu^\star$ yield the vector of efficient influence functions

$$\widetilde{\psi}_{P_\nu}^{(j)} = (A_\nu^\star)^{-1}\widetilde{\chi}_\nu^{(j)} = (A_\nu^\star)^{-1}\zeta^{(j)} - \int \zeta^{(j)}\mathrm{d}\nu, \quad j = 1, \dots, d,$$

and the assertion follows from Theorem 3.8. □

Remark 3.20. The assumption $\varphi_\varepsilon(u) \neq 0, u \in \mathbb{R}$, is not sufficient for the injectivity of $A_\nu^\star$ as the following counterexample shows: Let $\nu = \delta_0$ be the Dirac measure in zero and $\mu = \mathcal{N}(0,1)$ be standard normal such that $P_\nu = \nu * \mu = \mathcal{N}(0,1)$. Consider $g \in L_0^2(P_\nu)$ with $g(x) = x^3, x \in \mathbb{R}$. Then, $A_\nu^\star g(x) = \mathbb{E}[(x+\varepsilon_1)^3]$ is zero at the origin. Hence, $A_\nu^\star g = 0$ ν-a.s. and $0 \neq g \in \ker A_\nu^\star$. A sufficient condition for $A_\nu^\star$ being injective is given in Theorem 3.19 by assuming additionally a Lebesgue density of ν, which is a natural assumption. In particular, we obtain $\overline{\operatorname{ran}} A_\nu = L_0^2(P_\nu)$ implying that the tangent space $\dot{\mathcal{P}}_{P_\nu} = A_\nu \dot{\Theta}_\nu$ is dense in the set of all score functions. Without injectivity the convolution theorem remains true if the inverse $(A_\nu^\star)^{-1}$ in (3.20) is replaced by the Moore–Penrose pseudoinverse $(A_\nu^\star)^\dagger$.

In view of Theorem 3.8 our condition $\zeta^{(j)} \in \operatorname{ran} A_\nu^\star, j = 1, \dots, d$, is necessary and sufficient for the regularity of the parameter. The remaining question is under which conditions $\mathbb{1}_{(-\infty,t]} \in \operatorname{ran} A_\nu^\star, t \in \mathbb{R}$, and how the pre-image looks like in this case. Applying the approach by Nickl and Reiß (2012), we will find an answer in the next section.

3.2.2. Upper bound

In Söhl and Trabs (2012b) we have shown that the plug-in estimator for a large class of linear functionals, among them the distribution function estimator from Chapter 4, is regular and asymptotically normal distributed in a situation where the parametric rate can be attained. In this section we will reproduce these results and show that asymptotic distribution coincides with the lower bound in Theorem 3.19.

As later in Chapter 4 we assume more specifically the existence of probability densities f and f_ε of X_j and ε_j, respectively. Recall the definition of the empirical characteristic function $\varphi_n(u) = \frac{1}{n}\sum_{j=1}^n e^{iuY_j}$. Using the kernel density estimator from (4.12) we estimate the linear functional $\vartheta = \langle \zeta, f \rangle$, for ζ specified later, by

$$\widehat{\vartheta}_h := \int \zeta(x)\widehat{f}_h(x)\mathrm{d}x = \int \zeta(x)\,\mathcal{F}^{-1}\Big[\mathcal{F}\,K_h\frac{\varphi_n}{\varphi_\varepsilon}\Big](x)\mathrm{d}x, \qquad (3.21)$$

with kernel K, bandwidth $h > 0$ and the usual notation $K_h(x) = h^{-1}K(x/h)$. Choosing $\mathcal{F}K = \mathbb{1}_{[-U,U]}$ for some $U > 0$ leads to the estimator proposed by Butucea and Comte (2009). Throughout, we suppose:

Assumption 3.B.

(i) $K \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ is symmetric and band–limited with $\operatorname{supp}(\mathcal{F}K) \subseteq [-1, 1]$,

(ii) for $l = 1, \dots, \ell$

$$\int K(x)\mathrm{d}x = 1, \quad \int x^l K(x)\mathrm{d}x = 0, \quad \int |x^{\ell+1}K(x)|\mathrm{d}x < \infty \quad \text{and} \tag{3.22}$$

(iii) $K \in C^1(\mathbb{R})$ satisfies, denoting $\langle x\rangle := (1 + x^2)^{1/2}$,

$$|K(x)| + |K'(x)| \lesssim (1 + |x|)^{-2}. \tag{3.23}$$

In view of classical central limit theorems in a model without additive errors, where no assumptions on the smoothness of the distribution are needed, we want to assume as less smoothness of f as possible still guaranteeing $\sqrt{n}$-rates. For some $\delta > 0$ the following assumptions on the density f will be needed:

Assumption 3.C. ν admits a Lebesgue density f satisfying

(i) $f \in L^\infty(\mathbb{R})$ and $\int |x|^{2+\delta} f(x)\mathrm{d}x < \infty$,

(ii) $f \in H^\alpha(\mathbb{R})$ that is the density has Sobolev smoothness of order $\alpha \geqslant 0$.

The boundedness of the observation density f_Y follows immediately from (i) since $\|f_Y\|_\infty \leqslant \|f\|_\infty \|f_\varepsilon\|_{L^1} < \infty$. In addition to the smoothness of f, the smoothness of ζ will be crucial. Since we measure the regularity of f and ζ in the same scale, it is natural to use Sobolev spaces. We assume for $\gamma_s, \gamma_c > 0$

$$\zeta \in Z^{\gamma_s,\gamma_c} := \Big\{\zeta = \zeta^c + \zeta^s \Big| \zeta^s \in H^{\gamma_s}(\mathbb{R}) \text{ is compactly supported and } \tfrac{1}{ix+1}\zeta^c(x) \in H^{\gamma_c}(\mathbb{R}), \zeta^c \in L^\infty(\mathbb{R})\Big\}. \tag{3.24}$$

Let us give two examples for ζ and corresponding γ_s, γ_c.

Example 3.21. To estimate the distribution function of X_j, one has to consider translations of the indicator function $\mathbb{1}_{(-\infty,0]}(x)$, $x \in \mathbb{R}$. Let a be a monotone decreasing $C^\infty(\mathbb{R})$ function, which is equal to zero for all $x \geqslant 0$ and for some $M > 0$ equal to one for all $x \leqslant -M$. We define $\zeta^s := \mathbb{1}_{(-\infty,0]} - a$ and $\zeta^c := a$. From the bounded variation of ζ^s follows $\zeta^s \in B^1_{1,\infty}(\mathbb{R}) \subseteq H^{\gamma_s}(\mathbb{R})$ for any $\gamma_s < 1/2$ by Besov smoothness of bounded variation functions (A.10) as well as by the Besov space embeddings (A.6) and (A.7). Since $a \in C^\infty(\mathbb{R})$, the condition on ζ^c is satisfied for any $\gamma_c > 0$. Hence, $\mathbb{1}_{(-\infty,t]} \in Z^{\gamma_s,\gamma_c}$ if $\gamma_s < 1/2$. On the other hand, this cannot hold for $\gamma_s > 1/2$ since $H^{\gamma_s}(\mathbb{R}) \subseteq C^0(\mathbb{R})$ by Sobolev's embedding theorem.

Example 3.22. In the context of M-estimation (or Z-estimation) the root of the equation

$$\langle \zeta(\cdot - t), f \rangle = 0$$

is used for inference, e.g., on the location of the distribution of X_j. A popular example in robust statistics is the Huber estimator where $\zeta(x) = h_K(x) := ((-K) \vee x) \wedge K$ for some $K > 0$. In that case a similar decomposition as in Example 3.21 shows $h_K \in Z^{\gamma_s,\gamma_c}$ for $\gamma_s < 3/2$.

The ill-posedness of the problem is determined by the decay of the characteristic function of the errors. More precisely, we suppose

Assumption 3.D. Let the error distribution μ has a Lebesgue density f_ε satisfying

(i) $\int |x|^{2+\delta} f_\varepsilon(x)\mathrm{d}x < \infty$ thus φ_ε is twice continuously differentiable,

(ii) $|\varphi_\varepsilon(u)| \neq 0$ for all $u \in \mathbb{R}$ and

(iii) $|(\varphi_\varepsilon^{-1})'(u)| \lesssim (1+|u|)^{\beta-1}$ for some $\beta > 0$, in particular $|\varphi_\varepsilon^{-1}(u)| \lesssim (1+|u|)^{\beta}, u \in \mathbb{R}$.

Throughout, we write $\varphi_\varepsilon^{-1} = 1/\varphi_\varepsilon$. The Assumption (iii) on the distribution of the errors is similar to the classical decay assumption by Fan (1991a) and it is fulfilled for many ordinary smooth error laws such as gamma or Laplace distributions as discussed above. Assumption 3.D(iii) implies that φ_ε^{-1} is a Fourier multiplier on Besov spaces so that

$$B^s_{p,q}(\mathbb{R}) \ni f \mapsto \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F} f] \in B^{s-\beta}_{p,q}(\mathbb{R})$$

for $p, q \in [1, \infty], s \in \mathbb{R}$, is a continuous linear map, which is essential in our proofs, see Lemma 3.41. On Assumption 3.D and for $\zeta \in Z^{\gamma_s,\gamma_c}$ we can rigorously interpret the action of the deconvolution operator $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]$ on ζ: The product rule for differentiation yields

$$\begin{aligned}
&\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta^c(x) = \mathcal{F}^{-1}\left[\varphi_\varepsilon^{-1}(-u)(\mathrm{Id}+\mathrm{D})\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)\right](x)\\
&= \mathcal{F}^{-1}\left[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)\right](x) + \mathcal{F}^{-1}\left[\varphi_\varepsilon^{-1}(-u)\big(\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)]\big)'(u)\right](x)\\
&= (1+ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x) + \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x)
\end{aligned}$$

and thus

$$\begin{aligned}
\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(x) = &\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)](x)\\
&+ (1+ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x)\\
&+ \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x).
\end{aligned} \tag{3.25}$$

While $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta$ may exist only in distributional sense in general, it is defined rigorously through the right-hand side of the above display for $\zeta \in Z^{\gamma_s,\gamma_c}$. This indicates why we have imposed an assumption on $(\varphi_\varepsilon^{-1})'$ and have defined $\|\bullet\|_{Z^{\gamma_s,\gamma_c}}$ as above. Denoting by $\lfloor \alpha \rfloor$ the largest integer smaller or equal to α, we state the central limit theorem

Theorem 3.23. *Grant Assumptions 3.C with $\alpha > 0$ and 3.D with $\beta > 0$ as well as $\zeta^{(1)}, \dots, \zeta^{(d)} \in Z^{\gamma_s,\gamma_c}$ with $\gamma_s > \beta$ and $\gamma_c > (1/2 \vee \alpha) + \gamma_s$. Furthermore, let the kernel K satisfy Assumption 3.B with $\ell = \lfloor \alpha + \gamma_s \rfloor$. If $h_n^{\alpha+\gamma_s} = o(n^{-1/2})$, then the estimators $\widehat{\vartheta}_h^{(j)} := \langle \zeta^{(j)}, \widehat{f}_h \rangle, j = 1, \dots, d$, are regular with rate $n^{-1/2}$ and*

$$\sqrt{n}\Big(\langle \zeta^{(1)}, \widehat{f}_h - f \rangle, \dots, \langle \zeta^{(d)}, \widehat{f}_h - f \rangle\Big) \Rightarrow \mathbb{G} \quad \text{as } n \to \infty$$

where $\mathbb{G}$ is $\mathcal{N}(0, \Sigma)$-distributed with covariance

$$\Sigma_{j,k} := \int \Big(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta^{(j)}\Big)\Big(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta^{(k)}\Big) \mathrm{d}P_\nu - \langle \zeta^{(j)}, f \rangle \langle \zeta^{(k)}, f \rangle \quad (3.26)$$

for $j, k \in \{1, \dots, d\}$.

Remark 3.24. In Söhl and Trabs (2012b) we show that for ζ in a slightly smaller subset of Z^{γ_s,γ_c} the transition class $\{\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(\bullet - t) : t \in \mathbb{R}\}$ is P_ν–pregaussian and that the corresponding limit distribution $\mathbb{G}$ is tight in $\ell^\infty(\mathbb{R})$. Therefore, Theorem 3.23 can be extended to a Donsker theorem for transition classes in Z^{γ_s,γ_c}.

The proof of Theorem 3.23 is postponed to Section 3.4.1. To have $\sqrt{n}$-rates we suppose $\gamma_s > \beta$, which means that the smoothness of the functionals compensates the ill–posedness of the problem. This condition is natural in view of the abstract analysis in terms of Hilbert scales by Goldenshluger and Pereverzev (2003), who obtain the minimax rate $n^{-(\alpha+\gamma_s)/(2\alpha+2\beta)} \vee n^{-1/2}$ in our notation. As a consequence of the condition on γ_s and γ_c we can bound the stochastic error term of the estimator $\widehat{\vartheta}_h$ uniformly in $h \in (0, 1)$. The bias term is of order $h^{\alpha+\gamma_s}$. We illustrate the range of this theorem by the following examples.

Example 3.25. For estimating the distribution function Assumption 3.D needs to be fulfilled for some $\beta < 1/2$ owing to the condition $\gamma_s > \beta$. This is fulfilled by the gamma distribution $\Gamma(\beta, \eta)$ with $\beta \in (0, 1/2)$ and $\eta \in (0, \infty)$. This condition will be recovered in Corollary 4.9. For the Huber estimator from Example 3.22 we required $\beta < 3/2$, which holds, for instance, for the chi–squared distribution with one or two degrees of freedom or for the exponential distribution.

Example 3.26. Butucea and Comte (2009) studied the case $\beta > 1$ and derived $\sqrt{n}$-rates for $\gamma_s > \beta$ in our notation. In particular, they considered supersmooth ζ, that is $\mathcal{F}\zeta$ decays exponentially. In this case $\zeta \in H^s(\mathbb{R})$ for any $s \in \mathbb{N}$. Requiring the slightly stronger assumption that $\langle x \rangle^\tau \zeta(x) \in H^s(\mathbb{R})$ for some arbitrary small $\tau > 0$ and for all $s \in \mathbb{N}$ we can choose $\zeta^c := \zeta$ and $\zeta^s := 0$. Then β can be taken arbitrary large such that all gamma distributions, the Laplace distributions and convolutions of them can be chosen as error distributions.

Let us finally show that the asymptotic distribution in Theorem 3.23 coincides with the optimal distribution in the convolution Theorem 3.19.

Proposition 3.27. *In the deconvolution model* (3.13) *let $\nu \in \Theta$ have a bounded Lebesgue density and let μ satisfy Assumption 3.D for some $\beta > 0$. If $\zeta \in Z^{\gamma_s,\gamma_c}$ with $\gamma_s > \beta$ and $\gamma_c > (1/2 \vee \alpha) + \gamma_s$, then $\zeta \in \operatorname{ran} A_\nu^\star$ for $A_\nu^\star$ from* (3.19) *and it holds $(A_\nu^\star)^{-1}\zeta = \mathcal{F}^{-1}[1/\varphi_\varepsilon(-\bullet)] * \zeta$.*

Using that the deconvolution operator is a Fourier multiplier on Besov spaces, this proposition can be proved similarly to Proposition 3.39 below and thus the proof is omitted. In combination with Theorem 3.19. We can recover Theorem 4 in Söhl and Trabs (2012b) under weaker assumptions on the distributions of ν and μ:

Corollary 3.28. *In the deconvolution model* (3.13) *let $\nu \in \Theta$ have a bounded Lebesgue density and let μ satisfy Assumption 3.D for some $\beta > 0$. Let moreover $\zeta^{(1)}, \dots, \zeta^{(d)} \in Z^{\gamma_s,\gamma_c}$ with $\gamma_s > \beta$ and $\gamma_c > (1/2 \vee \alpha) + \gamma_s$. Then the limit distribution of any regular estimator of the parameter $(\int \zeta^{(1)} \mathrm{d}\nu, \dots, \int \zeta^{(d)} \mathrm{d}\nu)$ equals $\mathcal{N}(0,\Sigma) * M$ for some Borel probability measure M and with covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ given by* (3.26).

In particular, in the situation of Theorem 3.23 the plug-in estimator $\widehat{\vartheta}_h$ from (3.21) *is semiparametrically efficient.*

3.2.3. Efficient quantile estimation

Following the plug-in paradigm which we have outlined in Chapter 1, we define the quantile estimator for any $\tau \in (0,1)$

$$\widehat{q}_{\tau,h} := \operatorname{argmin}_{\eta \in [-U_n, U_n]} \Big| \int_{-\infty}^{\eta} \widehat{f}_h(x) \mathrm{d}x - \tau \Big|$$

with $\widehat{f}_h$ from (3.21) and $U_n \to \infty$ logarithmically fast. The generalization of $\widehat{q}_\tau$ to unknown error distributions will be comprehensively studied in Chapter 4. Using the previous insights, we can conclude that $\widehat{q}_{\tau,b}$ is semiparametrically efficient if it achieves the parametric rate. We restrict on the case $d = 1$ since the results for the multidimensional case can be analogously established.

For a given level τ and any probability measure $\nu \in \Theta$ we define the functional $\chi_\tau(\nu)$ as arbitrary point in $\mathbb{R}$ such that

$$\nu\big((-\infty, \chi_\tau(\nu)-]\big) \leqslant \tau \leqslant \nu\big((-\infty, \chi_\tau(\nu)]\big). \tag{3.27}$$

Since the precise choice of $\chi_\tau(\nu)$ is irrelevant, we can in particular choose the τ-quantile q_τ. To apply the convolution Theorem 3.8, we have to verify pathwise differentiability of χ_τ with respect to the tangent set $\dot{\Theta}_\nu$ from (3.14). It follows from the Hadamard-differentiability of the quantile function, cf. van der Vaart (1998, Chap. 21):

Lemma 3.29. *Let $\nu \in \Theta$ has a Lebesgue density f in a neighborhood of q_τ with $f(q_\tau) > 0$. Then χ_τ defined by* (3.27) *is pathwise differentiable with respect to $\dot{\Theta}_\nu = L^2_0(\nu)$ with derivative*

$$\dot{\chi}_{\tau,\nu} b = -\frac{\int_{-\infty}^{q_\tau} b \mathrm{d}\nu}{f(q_\tau)}, \quad b \in \dot{\Theta}_\nu.$$

Proof. For $b \in \dot{\Theta}_\nu$ we consider the paths $[0,\tau) \ni t \to \nu_t$ as defined in Section 3.2.1: ν_t is given by $\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu} = k(tb)/\int k(tb)\mathrm{d}\nu$ with the function k from Example 3.14. The assertion can be reformulated as

$$\Big|\frac{\chi_\tau(\nu + th_t) - \chi_\tau(\nu)}{t} - \dot{\chi}_{\tau,\nu} b\Big| \to 0 \quad \text{for } t \downarrow 0 \tag{3.28}$$

with $h_t = t^{-1}(\nu_t - \nu)$. With the calculations in the proof of Lemma 3.18 we see that dominated convergence yields

$$\limsup_{t\downarrow 0} \sup_{\eta\in\mathbb{R}} \frac{1}{t}\Big| \int_{-\infty}^{\eta} (\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu} - 1 - t\dot{\chi}_{\tau,\nu} b)\mathrm{d}\nu\Big| \to 0.$$

Since the functional χ_τ can be equivalently defined via the distribution functions of ν_t and because $\int_{-\infty}^{\eta} b\mathrm{d}\nu$ is continuous in $\eta \in \mathbb{R}$, Lemma 21.3 in van der Vaart (1998) shows Hadamard-differentiability of χ_τ which implies (3.28). □

Let us combine some basic findings of the following chapter with the central limit Theorem 3.23 to obtain the asymptotic distribution of $\widehat{q}_\tau$. Altogether we find the following result:

Theorem 3.30. *Let $\nu \in \Theta$ has the τ-quantile $q_\tau \in \mathbb{R}$ and a Lebesgue density $f \in C^{\alpha_0}([q_\tau - \zeta, q_\tau + \zeta])$ for some $\alpha_0, \zeta > 0$. Let f satisfy $f(q_\tau) > 0$ and Assumption 3.C with $\alpha > 0$. Grant Assumption 3.D with $\beta \in (0, 1/2)$ and Assumption 3.B with $\ell = \lfloor \alpha + \beta \rfloor$. If $\beta < \alpha$ and $h_n = n^{-1/(1+2\gamma)}$ for some $\gamma \in (\beta, \alpha)$, then the estimator $\widehat{q}_\tau$ is regular and semiparametrically efficient satisfying*

$$\sqrt{n}(\widehat{q}_{\tau,h} - q_\tau) \Rightarrow \mathcal{N}(0, \Sigma_\tau) \quad \textit{as } n \to \infty \tag{3.29}$$

with variance

$$\Sigma_\tau = \frac{1}{f^2(q_\tau)}\Big(\int \Big(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \mathbb{1}_{(-\infty,q_\tau]}\Big)^2 \mathrm{d}P_\nu - \nu((-\infty, q_\tau])^2\Big).$$

Proof. Note that all results in Section 4.4.1 carry over or even simplify to the case of known error distribution. Hence, we may apply the results from that section replacing $\widetilde{f}_h$ and $\widetilde{q}_{\tau,h}$ by $\widehat{f}_h$ and $\widehat{q}_{t,h}$, respectively. Applying the error representation (4.42) as well as (4.38) and (4.43), we obtain

$$\widehat{q}_{\tau,h} - q_\tau = -\frac{\int_{-\infty}^{q_\tau} (\widehat{f}_h(x) - f(x))\mathrm{d}x + o_P(n^{-1/2})}{f(q_\tau) + o_P(1)}. \tag{3.30}$$

Therefore, Theorem 3.23 and Slutsky's lemma yield (3.29).

The regularity of $\widehat{q}_{\tau,h}$ follows from the asymptotic linearity of $\int_{-\infty}^{q_\tau} \widehat{f}_h(x)\mathrm{d}x$ as shown in Section 3.4.1 and equation (3.30). To show the efficiency, we conclude from Lemma 3.29 that the gradient of χ_τ at ν is given by

$$\widetilde{\chi}_{\tau,\nu} = -\frac{\mathbb{1}_{(-\infty,q_\tau]} - \nu((-\infty, q_\tau])}{f(q_\tau)}.$$

The lower bound for the variance can then be proved analogously to Theorem 3.19 using Corollary 3.28. □

The asymptotic variance naturally corresponds to the classical case of direct observations. If ε_j are distributed according to the Dirac measure δ_0, the deconvolution operator degenerates to convolution with δ_0 and the variance would become $\tau(1-\tau)/f^2(q_\tau)$ which is exactly the asymptotic variance of the empirical quantile function, cf. Corollary 21.5 in van der Vaart (1998).

Let us briefly discuss the possibly surprising condition $\beta < \alpha$. From the error representation (4.9) we see that the bandwidth has to be chosen such that the density estimator is uniformly consistent which requires a minimal smoothing. This yields the condition $h_n^{1+2\beta}\frac{n}{\log n} \to \infty$. On the other hand the bias of the distribution function estimator must be negligible implying (roughly) $h_n^{1+2\alpha}n \to 0$. This trade-off is the reason for assumption $\beta < \alpha$. Note that α is the Sobolev regularity of f_ν. For $\alpha > 1/2$ the constraint is always satisfied, because $\beta < 1/2$, and the Sobolev embedding yields positive Hölder regularity similarly to Chapter 4 where this problem does not occur.

3.3. Lévy processes

3.3.1. Setting and upper bound

Let L be real-valued Lévy process with characteristic triplet (σ^2, γ, ν) as introduced in Section 2.1. We observe L at equidistant time points $t_k = \Delta k$ with $k = 0, \ldots, n$ for some fixed $\Delta > 0$ and for $n \to \infty$. Our aim is to derive a convolution theorem for the estimation of the linear functional of the jump measure

$$\chi(\nu) := \int \zeta \, \mathrm{d}\nu \quad \text{for } \zeta \in L^1(\nu) \cap L^2(\nu). \tag{3.31}$$

If ζ is $\mathbb{R}^d$-valued, the scalar product in (3.31) has to be interpreted coordinatewise. As a relevant example the reader should have in mind the generalized distribution function of ν introduced in the Chapter 1. It corresponds to $\zeta = \mathbb{1}_{(-\infty,t]}$ for $t < 0$ and $\zeta = \mathbb{1}_{[t,\infty)}$ for $t > 0$. In order for the estimation of $\chi(\nu)$ to be possible with parametric rate, we restrict on processes with finite variation in view of the lower bounds by Neumann and Reiß (2009). That means $\sigma^2 = 0$ and $\int_{\mathbb{R}}(|x| \wedge 1)\nu(\mathrm{d}x) < \infty$ are assumed.

Due to the stationary and independent increments of L, the random variables $Y_k := L_{\Delta k} - L_{\Delta(k-1)}, k = 1, \ldots, n$, are independent and identically distributed. Their characteristic function is given by the Lévy-Khintchine formula, see Proposition 2.3,

$$\varphi_\nu(u) = \mathbb{E}\left[e^{iuL_\Delta}\right] = e^{\Delta\psi(u)} \quad \text{with} \quad \psi(u) = i\gamma_0 u + \int (e^{iux} - 1)\nu(\mathrm{d}x), \quad u \in \mathbb{R}, \tag{3.32}$$

with $\gamma_0 = \gamma - \int_{-1}^{1} x\nu(\mathrm{d}x)$. Fixing the drift γ_0, the model is given by

$$\begin{aligned} &\mathcal{P} = \{P_\nu = \mathcal{F}^{-1}\varphi_\nu \,|\, \nu \in \Theta\} \quad \text{with} \\ &\Theta = \Big\{\nu \text{ jump measure on } (\mathbb{R}, \mathscr{B}(\mathbb{R}))\Big| \int (|x| \wedge 1)\nu(\mathrm{d}x) < \infty\Big\}. \end{aligned} \tag{3.33}$$

Before we apply the general results from Section 3.1 to this low-frequency Lévy model, we will give as a benchmark the asymptotic distribution of a potentially efficient estimator of the generalized distribution function. Due to the finite variation of the process L, we can use a simpler estimator than the one which will be studied in Section 5.2. Taking the first derivative of the characteristic exponent in (3.32) and assuming $\int |x|\nu(\mathrm{d}x) < \infty$, we obtain

$$\psi'(u) = \frac{\varphi_\nu'(u)}{\Delta\varphi_\nu(u)} = i\gamma_0 + i\,\mathcal{F}[x\nu](u).$$

With an analogous construction as in (5.5) we get the estimator proposed by Nickl and Reiß (2012). Using the empirical characteristic function of the increments $(Y_k)_{k=1,\dots,n}$, we find an empirical version $\widehat{\psi}_n'$ of ψ'. The distribution function estimator is then given by

$$\widehat{N}_h(t) = -\int g_t(x)\,\mathcal{F}^{-1}\Big[\widehat{\psi}_n'(u)\,\mathcal{F}\,K(hu)\Big](x)\mathrm{d}x \quad \text{with} \quad g_t(x) := \begin{cases} x^{-1}\mathbb{1}_{(-\infty,t]}, & t<0, \\ x^{-1}\mathbb{1}_{[t,\infty)}, & t>0, \end{cases} \tag{3.34}$$

for an appropriate band-limited kernel K and bandwidth $h > 0$. Under suitable assumptions, imposing especially $|\varphi_\nu^{-1}| \lesssim (1+|\bullet|)^\beta$ for some $\beta \in (0, 1/2)$, Nickl and Reiß (2012) show for $t_1, \dots, t_d \in \mathbb{R}$ that

$$\sqrt{n}\Big(\widehat{N}_h(t_1) - N(t_1), \dots, \widehat{N}_h(t_d) - N(t_d)\Big) \Rightarrow \mathbb{G} \quad \text{as } n \to \infty$$

where $\mathbb{G}$ is $\mathcal{N}(0,\Sigma)$-distributed with covariance

$$\Sigma_{j,k} := \Delta^{-2}\int \Big(\mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)] * (xg_{t_j}(x))\Big)\Big(\mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)] * (xg_{t_k}(x))\Big)\mathrm{d}P_\nu \tag{3.35}$$

for $j, k \in \{1, \dots, d\}$.

3.3.2. Regularity

Compared to tangents at the set of probability measures in Example 3.14, directions for the Lévy measures do not need to be centered since Lévy measures are not normalized. In general, jump measures are even not finite such that $L^2(\nu)$, which gives the Hilbert space structure, is still too large. We should intersect with $L^1(\nu)$ to include linear functionals as (3.31). Hence, we define the tangent space at $\nu \in \Theta$ as

$$\dot{\Theta}_\nu := L^1(\nu) \cap L^2(\nu) = \operatorname{lin} \dot{\Theta}_\nu. \tag{3.36}$$

Using the function $k(y) = 2/(1+e^{-2y})$ from Example 3.14, for any $b \in \dot{\Theta}_\nu$ the path $[0,1) \ni t \mapsto \nu_t$ with $\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) = k\big(tb(x)\big)$ is contained in Θ and satisfies

$$b(x) = \frac{\partial}{\partial t}\Big|_{t=0} \log\Big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x)\Big).$$

On this path the derivative of the functional (3.31) can be calculated with use of dominated convergence, noting that $|k(tb) - 1| \leqslant t|b| \in L^2(\nu)$. Hence,

$$\begin{aligned}\lim_{t\to 0} t^{-1}(\chi(\nu_t) - \chi(\nu)) &= \lim_{t\to 0}\int \zeta(x)t^{-1}\Big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) - 1\Big)\mathrm{d}\nu(x) \\ &= \int \zeta(x)b(x)\nu(\mathrm{d}x) = \langle \zeta, b\rangle_\nu =: \dot{\chi}_\nu b\end{aligned}$$

and thus the gradient is given by $\widetilde{\chi}_\nu = \zeta$. Compared to the deconvolution setting, we do not need to center $\widetilde{\chi}_\nu$ because the total mass of the Lévy measure is allowed to change along the path.

To apply Corollary 3.15, we need to verify Assumption 3.A for the Lévy model. By the Lévy–Khintchine representation the laws P_{ν_t} satisfy

$$\begin{aligned} P_{\nu_t} &= \mathcal{F}^{-1}\Big[\exp\Big(\Delta\Big(i\gamma_0 u + \int (e^{iux}-1)\nu_t(\mathrm{d}x)\Big)\Big)\Big] \\ &= \mathcal{F}^{-1}\Big[\exp\Big(\Delta\int (e^{iux}-1)\Big(k(tb(x))-1\Big)\nu(\mathrm{d}x)\Big)\varphi_\nu(u)\Big]. \end{aligned} \tag{3.37}$$

Owing to $(k(tb)-1)\in L^1(\nu)$, the measure P_{ν_t} is a convolution of P_ν and a compound Poisson type measure with signed jump measure $\Delta(k(tb)-1)\mathrm{d}\nu$. To see that the submodel $t\mapsto P_{\nu_t}$ is dominated, we check that the Hellinger distance of the jump measures $\int(\sqrt{\mathrm{d}\nu_t}-\sqrt{\mathrm{d}\nu})^2 = \int(\sqrt{k(tb(x))}-1)^2\nu(\mathrm{d}x)$ is finite for all t. Since the drift γ_0 remains constant, Theorem 33.1 by Sato (1999) yields absolute continuity $P_{\nu_t}\ll P_\nu$ for all t. To find the Hellinger derivative of the path $t\mapsto \frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu}$ at $t=0$, we note by dominated convergence

$$\begin{aligned} &\mathcal{F}^{-1}\Big[\frac{\partial}{\partial t}\Big|_{t=0}\exp\Big(\Delta\int(e^{iux}-1)(k(tb(x))-1)\nu(\mathrm{d}x)\Big)\varphi_\nu(u)\Big] \\ =&\Delta\,\mathcal{F}^{-1}\Big[\varphi_\nu(u)\Big(\int(e^{iux}-1)b(x)\nu(\mathrm{d}x)\Big)\Big] \\ =&\Delta\Big(P_\nu*(b\nu)-\int b(x)\nu(\mathrm{d}x)P_\nu\Big). \end{aligned}$$

This indicates how the score operator should look like. The following proposition determines the score operator A_ν and shows Hellinger regularity of the parametric submodel $t\mapsto P_{\nu_t}$. This is the key result to apply the theory of Section 3.1 to the Lévy model.

Proposition 3.31. *Let the model $\mathcal{P}$ be given by* (3.33) *with the tangent space $\dot\Theta_\nu$ at $\nu\in\Theta$ as defined in* (3.36)*. Then $P_\nu*(b\nu)\ll P_\nu$ for all $b\in\dot\Theta_\nu\cap L^\infty(\nu)$. Moreover, the linear operator*

$$A_\nu\big|_{\dot\Theta_\nu\cap L^\infty(\nu)}:\dot\Theta_\nu\cap L^\infty(\nu)\to L_0^2(P_\nu),\quad b\mapsto\Delta\frac{\mathrm{d}(P_\nu*(b\nu))-(\int b\mathrm{d}\nu)\mathrm{d}P_\nu}{\mathrm{d}P_\nu} \tag{3.38}$$

is bounded. $\dot\Theta_\nu\cap L^\infty(\nu)$ is dense in $\dot\Theta_\vartheta$ and thus $A_\nu:\dot\Theta_\nu\to L_0^2(P_\nu)$ can be defined as its unique continuous extension. Then for all $b\in\dot\Theta_\nu$ the associated submodel $[0,1)\ni t\mapsto P_{\nu_t}$ is Hellinger differentiable at zero with derivative $A_\nu b$, that means (3.10) *is fulfilled.*

The proof of this proposition is given in Section 3.4.2. An essential ingredient is an estimate of the Hellinger integral of two infinitely divisible distributions by Liese (1987). More precisely, his results imply (for details see Section 3.4.2)

$$\int\Big(\frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu}\Big)^2\mathrm{d}P_\nu\leqslant\exp\Big(\frac{1}{2}\int\big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}-1\big)^2\mathrm{d}\nu\Big)\leqslant\exp\Big(\frac{1}{2}t^2\|b\|_{L^2(\nu)}^2\Big). \tag{3.39}$$

Proposition 3.31 shows that the Lévy model $\mathcal{P}$, defined in (3.33) equipped with the tangent space $\dot\Theta_\nu$ from (3.36) satisfies Assumption 3.A. In particular, it is a regular indirect model at any $\nu \in \Theta$. Having in mind the regularity Lemma 3.18 in the deconvolution model, the score operators look very similar in both models. Since the gradient does not have to integrate to zero in the Lévy model, the centering is incorporated in the operator A_ν. Apart from that the convolution structure is the same. Therefore, the Lévy model can locally be weakly approximated with a linear white noise model whose operator is of deconvolution type.

By Proposition 3.31 the score operator A_ν is characterized by (3.38). To prove information bounds, we will combine this result with Proposition 3.12 which shows that it is sufficient to study $A_\nu^\star$ on a nicely chosen, dense subset of $L_0^2(P_\nu)$. Then Theorem 3.8 provides the convolution theorem for all $\zeta \in \operatorname{ran} A_\vartheta^\star$. In the following we will discuss Lévy processes with finite and infinite jump activity separately because the analytical properties of the score operator are quite different: In the compound Poisson case the inverse adjoint score operator can be explicitly expressed as a convolution with a finite signed measure. If the jump intensity is infinite, the distribution P_ν possesses a Lebesgue density and thus $A_\nu^\star$ will be a smoothing operator.

3.3.3. Compound Poisson processes

Let L be a compound Poisson process with jump intensity $\lambda := \nu(\mathbb{R}) < \infty$. Consequently, the tangent space simplifies to $\dot{\Theta}_\vartheta = L^2(\nu)$ and the measure P_ν can be written as the convolution exponential (cf. Sato, 1999, Rem. 27.3)

$$P_\nu = \delta_{\Delta\gamma_0} * \Big(e^{-\Delta\lambda} \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} \nu^{*k}\Big). \tag{3.40}$$

Define the subsets

$$\begin{aligned} \mathcal{G} &:= L^\infty(P_\nu) \cap L_0^2(P_\nu) \subseteq L_0^2(P_\nu) \quad \text{and} \\ \mathcal{H} &:= \{h \in L^2(\nu) | \sup_{k=0,1,\ldots} \|h\|_{L^\infty(\nu^{*k})} < \infty\} \subseteq \dot{\Theta}_\nu, \end{aligned} \tag{3.41}$$

which are dense in $L_0^2(P_\nu)$ and $L^2(\nu)$, respectively. Let $g \in \mathcal{G}$ and $b \in \mathcal{H}$. By Proposition 3.31 we know $(b\nu) * P_\nu \ll P_\nu$, which implies $g \in L^\infty((b\nu) * P_\nu)$. Hence, $\int g \mathrm{d}P_\nu = 0$ and Fubini's theorem yield

$$\begin{aligned} \langle A_\nu b, g\rangle_{P_\nu} = \int (A_\nu b) g \mathrm{d}P_\nu &= \Delta \int g \mathrm{d}(P_\nu * (b\nu)) - \Delta\Big(\int b \mathrm{d}\nu\Big)\Big(\int g \mathrm{d}P_\nu\Big) \\ &= \Delta \int\int g(x+y) b(x) \mathrm{d}\nu(x) \mathrm{d}P_\nu(y) \\ &= \Delta \langle P_\nu(-\bullet) * g, b\rangle_\nu. \end{aligned} \tag{3.42}$$

Therefore, the adjoint score operator on $\mathcal{G}$ is $A_\nu^\star|_{\mathcal{G}} \colon \mathcal{G} \to L^2(\nu), g \mapsto \Delta P_\nu(-\bullet) * g$.

Lemma 3.32. *The map $A_\nu^\star|_{\mathcal{G}} \colon \mathcal{G} \to \mathcal{H}, g \mapsto P_\nu(-\bullet) * g$ is well defined.*

Proof. First, we show $\nu^{*l} * P_\nu \ll P_\nu$ for any $l \in \mathbb{N}$. Let $A \in \mathscr{B}(\mathbb{R})$ satisfy $P_\nu(A) = 0$. (3.40) yields

$$0 = e^{-\Delta\lambda} \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} \int \mathbb{1}_A(x + \Delta\gamma_0) \mathrm{d}\nu^{*k}(\mathrm{d}x)$$

and thus $\int \mathbb{1}_A(x+\Delta\gamma_0)\mathrm{d}\nu^{*k}(\mathrm{d}x) = 0$ for all $k \in \mathbb{N}$. But this implies by linearity of the convolution that $\nu^{*l} * P_\nu(A) = 0$.

To see that $A_\nu^\star$ is well defined on equivalence classes with respect to P_ν zero sets, note that $\nu * P_\nu \ll P_\nu$ implies that for any function g with $g(x) = 0$ for P_ν-a.e. $x \in \mathbb{R}$ it holds $P_\nu(-\bullet) * g(y) = 0$ for ν-a.e. $y \in \mathbb{R}$. It remains to show $A_\nu^\star g \in \mathcal{H}$ for $g \in \mathcal{G}$. For any $g \in L^\infty(P_\nu)$ there is a set $A \in \mathscr{B}(\mathbb{R})$ with $P_\nu(A) = 0$ such that $g(y) \leqslant C$ for some constant $C > 0$ and for all $y \notin A$. Using

$$0 = \nu^{*l} * P_\nu(A) = \int\int \mathbb{1}_{A-\{x\}}(y) P_\nu(\mathrm{d}y)\nu^{*l}(\mathrm{d}x)$$

we infer $P_\nu(A-\{x\}) = 0$ for ν^{*l}-a.e. $x \in \mathbb{R}$ and therefore $P_\nu(-\bullet)*g(y) = \int g(x+y)P_\nu(\mathrm{d}x)$ is bounded by C for ν^{*l}-a.e. $y \in \mathbb{R}$. Hence, $\|P_\nu(-\bullet) * g\|_{L^\infty(\nu^{*l})} \leqslant C$ for any $l \in \mathbb{N}$. □

Although the centering of $g \in L_0^2(P_\nu)$ implies $A_\nu^\star g(0) = \int g\mathrm{d}P_\nu = 0$, it does not cause an additional constraint owing to $\nu(\{0\}) = 0$. In general, $A_\nu^\star$ is not injective as the following example shows:

Example 3.33 (Poisson process)**.** Setting $\nu = \delta_1, \gamma_0 = 0$ and $\Delta = \lambda = 1$, the law $P_\nu = e^{-1}\sum_{k=0}^\infty \delta_k/(k!)$ is the Poisson distribution and the adjoint score operator is given by

$$A_\nu^\star g(x) = e^{-1}\sum_{k=0}^\infty g(x+k)/(k!), \quad x \in \mathbb{R}\,.$$

Consider the function $g = \mathbb{1}_{\{0\}} - 2\mathbb{1}_{\{1\}} + 2\mathbb{1}_{\{2\}}$ which is a nonzero element of $\mathcal{G}$ by construction. However, $A_\nu^\star g(1) = 0$ and thus $0 \neq g \in \ker A_\nu^\star$ contradicting injectivity.

As in the deconvolution model we assume therefore that ν admits a Lebesgue density concluding injectivity of $A_\nu^\star$ exactly as in Theorem 3.19. Since (3.32) yields $|\varphi_\nu(u)| = e^{\Delta\int(\cos(ux)-1)\nu(x)\mathrm{d}x} \geqslant e^{-2\Delta\lambda}$ for all $u \in \mathbb{R}$, the inverse of $A_\nu^\star$ is then the deconvolution operator $h \mapsto \Delta^{-1}\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * h$ with the finite signed measure

$$\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] = \delta_{-\Delta\gamma_0} * \Big(e^{\Delta\lambda}\sum_{k=0}^\infty \frac{(-\Delta)^k}{k!}\nu(-\bullet)^{*k}\Big),$$

which is well defined on $\mathcal{H}$. In particular, the pre-image of the indicator function $\zeta = \mathbb{1}_{(-\infty,t]}$ (or equivalently $\mathbb{1}_{(-\infty,t]}\mathbb{1}_{\mathbb{R}\setminus\{0\}}$) is well defined for any $t \in \mathbb{R}$. Consequently, Corollary 3.15 and Theorem 3.8 yield

Corollary 3.34. *Let L be a pure jump process of compound Poisson type with jump measure ν which is absolutely continuous with respect to the Lebesgue measure. Then the limit distribution of any regular estimator of the distribution function $\mathbb{R}^d \ni (t_1,\dots,t_d) \mapsto (\nu((-\infty,t_1]),\dots,\nu((-\infty,t_d]))$, for $d \in \mathbb{N}$, is a convolution $\mathcal{N}(0,\Sigma) * M$ for some Borel probability measure M and with covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ given by*

$$\Sigma_{i,j} = \Delta^{-2}\int\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \mathbb{1}_{(-\infty,t_i]}\Big)\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \mathbb{1}_{(-\infty,t_j]}\Big)\mathrm{d}P_\nu$$

for $i,j \in \{1,\dots,d\}$.

Considering the negative half line, this lower bound coincides with the asymptotic variance (3.35) of the kernel estimator by Nickl and Reiß (2012). An interesting deviation is obtained by restricting the model on compound Poisson processes with fixed jump intensity $\lambda > 0$ as studied in the *decompounding* problem by Buchmann and Grübel (2003). Similarly to the deconvolution model in Section 3.2 the tangent space is then given by $L_0^2(\nu)$ and thus the gradient of the functional $\chi(\nu) = \nu((-\infty,t])$ equals $\widetilde{\chi}_\nu = \mathbb{1}_{(-\infty,t]} - \nu((-\infty,t])$. We obtain the smaller information bound, setting $d = 1$ for simplicity,

$$\Delta^{-2}\int\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big]*\mathbb{1}_{(-\infty,t]}\Big)^2\mathrm{d}P_\nu - \Delta^{-2}\nu((-\infty,t])^2.$$

That means an efficient estimator which "knows" the jump intensity should have a smaller variance than for unknown λ and the statistical problem is significantly simpler. Indeed, the estimator by Buchmann and Grübel (2003) is asymptotically normal with the above variance.

3.3.4. Lévy processes with infinite jump activity

If the Lévy process has infinite jump activity, the analysis is more difficult. However, we can profit from the absolute continuity of the infinite divisible distribution P_ν with respect to the Lebesgue measure (Sato, 1999, Thm. 27.4). To apply Fourier methods, we will again assume that ν admits a Lebesgue density which implies in particular that the set of Lebesgue-a.e. equivalence classes embeds into the ν-a.e. and into the P_ν-a.e. equivalence classes. Keeping the Hilbert space structure, we can then define

$$\mathcal{G} := H^\infty(\mathbb{R}) \cap L_0^2(P_\nu) \quad \text{and} \quad \mathcal{H} := \{b \in H^\infty(\mathbb{R}) | b(0) = 0\} \cap \dot{\Theta}_\nu, \tag{3.43}$$

where $H^\infty(\mathbb{R}) := \bigcap_{s\geqslant 0} H^s(\mathbb{R})$ with Sobolev spaces $H^s(\mathbb{R})$ of regularity $s \geqslant 0$ defined in (A.1). For $b \in \mathcal{H}$ the condition $b(0) = 0$ should hold for the continuous version of b. To allow that the generalized distribution function of ν can be estimated with parametric rate, we concentrate on mildly ill-posed problems leading to the assumption that $|\varphi_\nu(u)|$ decays polynomially as $|u| \to \infty$ (cf. Neumann and Reiß, 2009).

Lemma 3.35. *Let the finite variation Lévy process L with $\nu \in \Theta$ have infinite jump activity satisfying $|\varphi_\nu(u)| \gtrsim (1+|u|)^{-\beta}$ for some $\beta > 0$ and let ν be absolutely continuous with respect to the Lebesgue measure. Then*

(i) on $\mathcal{G}$ from (3.43) the adjoint score operator $A_\nu^\star|_\mathcal{G}$ is a bijection onto $\mathcal{H}$ satisfying

$$A_\nu^\star\big|_\mathcal{G}: \quad \mathcal{G} \to \mathcal{H} \subseteq \dot{\Theta}_\nu, \qquad g \mapsto \Delta(P_\nu(-\bullet) * g), \tag{3.44}$$

$$(A_\nu^\star)^{-1}\big|_\mathcal{H}: \quad \mathcal{H} \to \mathcal{G} \subseteq L_0^2(P_\nu), \qquad b \mapsto \Delta^{-1}\mathcal{F}^{-1}\big[\mathcal{F}b/\varphi_\nu(-\bullet)\big], \tag{3.45}$$

(ii) $\mathcal{G}$ is dense in $L_0^2(P_\nu)$ and $A_\nu^\star$ is the unique continuous extension of $A_\nu^\star|_\mathcal{G}$.

The proof of this lemma uses that by the polynomial decay $(1+|u|)^{-\beta} \lesssim |\varphi_\nu(u)| \lesssim 1$ both operators, $A_\nu^\star$ and $(A_\nu^\star)^{-1}$, are Fourier multipliers on Sobolev spaces and thus $H^\infty(\mathbb{R})$-functions are mapped into $H^\infty(\mathbb{R})$ again. The smoothness will be used to show

$|(A_\nu^\star g)(x)| \lesssim 1 \wedge |x|$ for $x \in \mathbb{R}$ in order to verify that $A_\nu^\star$ is well defined. For details we refer to Section 3.4.3.

Comparing the adjoint score operator (3.45) in the Lévy model to $A_\nu^\star$ in the deconvolution model (3.15), we see that both operators have exactly the same structure. To invert $A_\nu^\star$, we have to deconvolve with the observation measure itself in the Lévy case. From our lower bounds perspective we clearly recover this *auto-deconvolution* phenomenon, which was already described by Belomestny and Reiß (2006a) as well as Nickl and Reiß (2012). For convenience we will write throughout $\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)]*b = \mathcal{F}^{-1}[\mathcal{F}\, b/\varphi_\nu(-\bullet)]$, which is justified in distributional sense. In combination with the results for the compound Poisson case Lemma 3.35 has two immediate consequences.

Remark 3.36. If the Lévy process is of finite variation, has an absolutely continuous jump measure and has either finite jump activity or has a polynomial decreasing characteristic function, then

(i) $A_\nu^\star$ is injective and therefore $\overline{\operatorname{ran}} A_\nu = L_0^2(P_\nu)$. This means that the tangent set $\dot{\mathcal{P}}_{P_\nu} = A_\nu \dot{\Theta}_\nu$ is dense in $L^2(P_\nu)$.

(ii) For any linear functional $\chi(\nu) = \int \zeta \mathrm{d}\nu$ satisfying $\zeta \in \mathcal{H}$, where $\mathcal{H} \subseteq \dot{\Theta}_\nu$ is defined in (3.41) and (3.43), respectively, the information bound is given by

$$\Delta^{-2} \int \Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \zeta \Big)^2 \mathrm{d}P_\nu.$$

The subset $\mathcal{H}$ of arbitrary large Sobolev smoothness is obviously very restrictive. Let us extend the information bound to a larger class of functionals by using Proposition 3.12. This is illustrated in the following example.

Example 3.37 (Gamma process)**.** Let L be a gamma process with $Y_k \sim \Gamma(\alpha\Delta, \lambda)$ for all $k = 1, \dots, n$. For simplicity set $\lambda = 1$. The probability density, the characteristic function and the Lévy measure are given by

$$\gamma_{\alpha\Delta}(x) := \frac{1}{\Gamma(\alpha\Delta)} x^{\alpha\Delta-1} e^{-x} \mathbb{1}_{[0,\infty)}(x), \qquad \varphi_\nu(u) = (1-iu)^{-\alpha\Delta} \quad \text{and}$$
$$\nu(\mathrm{d}x) = \alpha x^{-1} e^{-x} \mathbb{1}_{[0,\infty)}(x)\mathrm{d}x, \quad \text{for } x, u \in \mathbb{R},$$

respectively. Therefore, $|\varphi_\nu|$ decays with polynomial rate $\beta = \Delta\alpha$ and we can apply Lemma 3.35. The estimation of the generalized distribution function $\chi(\nu) = \int_t^\infty \mathrm{d}\nu$ for some fixed $t > 0$, induces the gradient $\widetilde{\chi}_\nu = \mathbb{1}_{[t,\infty)}$. To approximate $\widetilde{\chi}_\nu$ with a sequence in $\mathcal{H}$, we construct $\chi_n(x) = \int_{-\infty}^x (\delta_n(y-t) - \delta_n(y-n))\mathrm{d}y$ for a Dirac sequence (δ_n). More precisely, let $(\delta_n) \subseteq C^\infty(\mathbb{R})$ be a family of smooth nonnegative functions satisfying $\int_{\mathbb{R}} \delta_n = 1$ and $\operatorname{supp} \delta_n \subseteq [-1/n, 1/n]$. Obviously, $(\chi_n) \subseteq \mathcal{H}$. Since ν is a finite measure on $\mathbb{R} \setminus (-\varepsilon, \varepsilon)$ for any $\varepsilon > 0$, dominated convergence shows $\|\widetilde{\chi}_\nu - \chi_n\|_{L^2(\nu)} \to 0$. Denoting the distribution function of $\Gamma(\beta,1)$ by Γ_β, we obtain for $\alpha\Delta<1/2$

$$\begin{aligned}(A_\nu^\star)^{-1}\chi_n &= \Delta^{-1}\,\mathcal{F}^{-1}[(1-iu)(1+iu)^{\alpha\Delta-1}]*\chi_n\\ &= \Delta^{-1}\gamma_{1-\alpha\Delta}(-\bullet)*(\chi_n-\chi_n')\\ &= \Delta^{-1}\big(\gamma_{1-\alpha\Delta}(-\bullet)*\chi_n-\gamma_{1-\alpha\Delta}(-\bullet)*\delta_n(\bullet-t)\\ &\qquad+\gamma_{1-\alpha\Delta}(-\bullet)*\delta_n(\bullet-n)\big)\\ &\to \Delta^{-1}\big((1-\Gamma_{1-\alpha\Delta}(t-\bullet))-\gamma_{1-\alpha\Delta}(t-\bullet)\big)\\ &= \Delta^{-1}\gamma_{1-\alpha\Delta}(-\bullet)*(\mathbb{1}_{(-\infty,t]}-\delta_t)=:\psi \end{aligned}\tag{3.46}$$

where the convergence holds in $L^2(\mathbb{R})$ owing to $\gamma_{1-\alpha\Delta}\in L^2(\mathbb{R})$ for $\alpha\Delta<1/2$. Therefore, in a natural way the limiting object is $\psi=\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)]*\mathbb{1}_{(-\infty,t]}$. When does this limit hold in $L^2(P_\nu)$, too? As we saw above, the probability density of P_ν is bounded everywhere except for the singularity at zero which is of order $1-\alpha\Delta$. For any $t>0$ and n large enough $\gamma_{1-\alpha\Delta}*\delta_n(\bullet-t)$ is uniformly bounded in a small neighborhood of zero such that dominated convergence around zero together with the $L^2(\mathbb{R})$-convergence on the real line yields $\|(A_\nu^\star)^{-1}\chi_n-\psi\|_{L^2(P_\nu)}\to 0$. Hence, Proposition 3.12 shows $\mathbb{1}_{[t,\infty)}\in\operatorname{ran}A_\nu^\star$. Therefore, the information bound is given by

$$\int(\mathcal{F}[1/\varphi_\nu(-\bullet)]*\mathbb{1}_{[t,\infty)})^2\mathrm{d}P_\nu,$$

which can be understood via (3.46) or equivalently as $\lim_{n\to\infty}\|(A_\nu^\star)^{-1}\chi_n\|^2_{L^2(P_\nu)}$. For $\alpha\Delta>1/2$ Neumann and Reiß (2009) show that $\nu([t,\infty))$ cannot be estimated with $\sqrt{n}$-rate.

This example shows the importance of the pseudo-locality for the devolution operator which was discussed by Nickl and Reiß (2012) in detail: If the singularity of the pointwise limit ψ as in (3.46) and the singularity of the distribution P_ν would coincide, the $L^2(P_\nu)$-norm of any approximating sequence $(A_\nu^\star)^{-1}\chi_n$ would diverge such that $\widetilde{\chi}_\nu$ cannot be an element of $\operatorname{ran}A_\nu^\star$ by Proposition 3.12. A simple example is given by the convolution of a Gamma process and a Poisson process (cf. Nickl and Reiß, 2012, Sect. 3.2).

Similarly to the deconvolution model in Section 3.2 we will show that for suitable regularity $\delta>0$ the class

$$Z^\delta(\mathbb{R}):=\Big\{\zeta=\zeta^s+\zeta^c\Big|\zeta^s,\frac{\zeta^s(x)}{x}\in H^\delta(\mathbb{R}),\zeta^c\in C^\delta(\mathbb{R}),\zeta^c(0)=0\Big\}$$

intersected with $L^1(\nu)\cap L^2(\nu)$ is a subset of $\operatorname{ran}A_\nu^\star$.

Example 3.38 (Generalized distribution function)**.** Recall that the generalized distribution function of ν corresponds to the functionals $\zeta_t:=\mathbb{1}_{(-\infty,t]}$ for $t<0$ and $\zeta_t:=\mathbb{1}_{[t,\infty)}$ for $t>0$. It is easy to check that ζ_t can be decomposed for all $t\neq 0$ in a way such that it is contained in $Z^\delta(\mathbb{R})$ for any $\delta<1/2$. For instance, write $\mathbb{1}_{[t,\infty)}=\zeta_t^s+\zeta_t^c$ with $\zeta_t^s(x):=e^{t-x}\mathbb{1}_{[t,\infty)}(x)$ and $\zeta_t^c(x):=(1-e^{t-x})\mathbb{1}_{[t,\infty)}(x)$ for $t>0$. Then ζ_t^s is a translation of the gamma density γ_1 such that its Fourier transform decays with polynomial rate one. The factor x^{-1} is harmless since ζ_t^s equals zero around the origin. Moreover, ζ_t^c is Lipschitz continuous. On the negative half line an analogous decomposition applies.

For this analytic description of the range of the adjoint score operator, let us apply the Fourier multiplier Theorem 5.5, which will be proved later, and the characterization in Proposition 3.12. We suppose that the Lévy process L with jump density ν satisfies the following assumption, which implies in particular a polynomial decay of $|\varphi_\nu|$ by Lemma 5.2.

Assumption 3.E. Let $x\nu$ admit a Lebesgue density $k \in BV(\mathbb{R})$ with $\beta := k(0+) + k(0-)$.

Proposition 3.39. *Let the finite variation Lévy process L with $\nu \in \Theta$ satisfy Assumption 3.E for some $\beta \geqslant 0$. If $\zeta \in Z^{\beta^+}(\mathbb{R}) \cap L^1(\nu) \cap L^2(\nu)$ for some $\beta^+ > \beta$, then $\zeta \in \operatorname{ran} A_\nu^\star$ holds with $(A_\nu^\star)^{-1}\zeta = \mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * \zeta$.*

The proof is postponed to Section 3.4.4. The formula $(A_\nu^\star)^{-1}\zeta = \mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * \zeta$ can be either understood in distributional sense or as the limit of an approximating sequence as illustrated in Example 3.37. Applying Theorem 3.8 and Proposition 3.39 on Example 3.38, we get the following convolution theorem.

Theorem 3.40. *Let the finite variation Lévy process L with $\nu \in \Theta$ satisfy Assumption 3.E for some $\beta < 1/2$. Then the limit distribution of any regular estimator of the generalized distribution function $(\mathbb{R} \setminus \{0\})^d \ni (t_1, \dots, t_d) \mapsto (\langle \zeta_{t_1}, \nu\rangle, \dots, \langle \zeta_{t_d}, \nu\rangle)$ is a convolution $\mathcal{N}(0, \Sigma) * M$ for some Borel probability measure M and with the covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ given by*

$$\Sigma_{i,j} = \frac{1}{\Delta^2}\int_{\mathbb{R}} \Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \zeta_{t_i}\Big)\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \zeta_{t_j}\Big)\mathrm{d}P_\nu$$

for $i, j \in \{1, \dots, d\}$.

In the situations of Corollary 3.34 and Theorem 3.40 the estimator (3.34) is therefore efficient. Analogously to the deconvolution model, these results can be extended to quantile estimation in the Lévy model, but we will not go into details.

3.4. Remaining proofs

3.4.1. Proof of Theorem 3.23

First, we provide an auxiliary lemma, which describes the properties of the deconvolution operator $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$.

Lemma 3.41. *Grant Assumptions 3.B and 3.D.*

(i) For all $s \in \mathbb{R}, p, q \in [1, \infty]$ the deconvolution operator $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]$ is a Fourier multiplier from $B^s_{p,q}(\mathbb{R})$ to $B^{s-\beta}_{p,q}(\mathbb{R})$, that is the linear map

$$B^s_{p,q}(\mathbb{R}) \to B^{s-\beta}_{p,q}(\mathbb{R}), f \mapsto \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F} f]$$

is bounded.

(ii) Let $\beta^+ > \beta$ and $f, g \in H^{\beta^+}(\mathbb{R})$. Then

$$\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f)g = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * g)f. \tag{3.47}$$

Using the kernel K, this equality extends to functions $g \in L^2(\mathbb{R}) \cup L^\infty(\mathbb{R})$ and finite Borel measures μ:

$$\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h] * \mu)g = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * g)\mathrm{d}\mu. \tag{3.48}$$

Proof.

(i) Analogously to Nickl and Reiß (2012), we deduce from Corollary 4.11 of Girardi and Weis (2003) that $(1+iu)^{-\beta}\varphi_\varepsilon^{-1}(-u)$ is a Fourier multiplier on $B^s_{p,q}$ by Assumption 3.D(iii). It remains to note that $j : B^s_{p,q}(\mathbb{R}) \to B^{s-\beta}_{p,q}(\mathbb{R}), f \mapsto \mathcal{F}^{-1}[(1+iu)^\beta \mathcal{F} f]$ is a linear isomorphism (Triebel, 2010, Thm. 2.3.8).

(ii) For $f \in H^{\beta^+}(\mathbb{R})$ (i) and the Besov embeddings (A.4), (A.6) and (A.7) yield

$$\| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f\|_{L^2} \lesssim \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f\|_{B^0_{2,1}} \lesssim \|f\|_{B^\beta_{2,1}} \lesssim \|f\|_{H^{\beta^+}} < \infty.$$

Therefore, it follows by Plancherel's equality

$$\begin{aligned}\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f)(x)g(x)\mathrm{d}x &= \frac{1}{2\pi}\int \varphi_\varepsilon^{-1}(-u)\mathcal{F} f(-u)\mathcal{F} g(u)\mathrm{d}u \\ &= \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * g)(x)f(x)\mathrm{d}x.\end{aligned}$$

To prove the second part of the claim for $g \in L^2(\mathbb{R})$, we note that by Young's inequality

$$\| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h]\|_{L^2} \leqslant \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\mathbf{1}_{[-1/h,1/h]}]\|_{L^2}\|K_h\|_{L^1} < \infty$$

due to the support of $\mathcal{F} K$ and Assumption (3.23) on the decay of K. Since μ is a finite measure and g is bounded, Fubini's theorem yields then

$$\begin{aligned}&\int g(x)(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h] * \mu)(x)\mathrm{d}x \\ &= \int\int g(x)\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h](x-y)\mu(\mathrm{d}y)\mathrm{d}x \\ &= \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * g)(y)\mu(\mathrm{d}y),\end{aligned}$$

where we have used the symmetry of the kernel. In order to apply Fubini's theorem for $g \in L^\infty(\mathbb{R})$, too, we have to show that $\| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h]\|_{L^1}$ is finite. We replace the indicator function by a function $\chi \in C^\infty(\mathbb{R})$ which equals one on $[-1/h, 1/h]$ and has got compact support. We estimate

$$\| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h]\|_{L^1} \leqslant \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\chi]\|_{L^1}\|K_h\|_{L^1}. \tag{3.49}$$

Using that $\varphi_\varepsilon^{-1}\chi$ is twice continuously differentiable and has compact support, we obtain

$$\begin{aligned}\|(1+x^2)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\chi](x)\|_\infty &\leqslant \|\,\mathcal{F}^{-1}[(\mathrm{Id}-\mathrm{D}^2)\varphi_\varepsilon^{-1}\chi](x)\|_\infty\\ &\leqslant \|(\mathrm{Id}-\mathrm{D}^2)\varphi_\varepsilon^{-1}\chi\|_{L^1}<\infty,\end{aligned}$$

where we denote the identity and the differential operator by Id and D, respectively. This shows that (3.49) is finite. □

To prove Theorem 3.23, we show first the asymptotic normality and conclude in a second step the regularity of the estimator.

Step 1: Since the class Z^{γ_s,γ_c} is an $\mathbb{R}$ vector space, the Cramér-Wold device yields that it suffices to consider $d=1$. As usual, we decompose the error into a stochastic error term and a bias term:

$$\widehat{\vartheta}_h-\vartheta=\int\zeta(x)\,\mathcal{F}^{-1}\Big[\mathcal{F}K_h\frac{\varphi_n-\varphi}{\varphi_\varepsilon}\Big](x)\mathrm{d}x+\int\zeta(x)(K_h*f-f)(x)\mathrm{d}x. \tag{3.50}$$

The bias term can be estimated by the standard kernel estimator argument. Let us consider the singular and the continuous part of ζ separately. Applying Plancherel's identity and Hölder's inequality, we obtain

$$\begin{aligned}&\int|\zeta^s(x)(K_h*f(x)-f(x))|\mathrm{d}x\\ &=\frac{1}{2\pi}\int|\,\mathcal{F}\zeta^s(u)(\mathcal{F}K(hu)-1)\,\mathcal{F}f(-u)|\mathrm{d}u\\ &\leqslant\|(1+|u|)^{-(\alpha+\gamma_s)}(\mathcal{F}K(hu)-1)\|_\infty\int(1+|u|)^{\alpha+\gamma_s}|\,\mathcal{F}\zeta^s(u)\,\mathcal{F}f(u)|\mathrm{d}u\\ &\leqslant h^{\alpha+\gamma_s}\|u^{-(\alpha+\gamma_s)}(\mathcal{F}K(u)-1)\|_\infty\|\zeta^s\|_{H^{\gamma_s}}\|f\|_{H^\alpha}.\end{aligned}$$

The term $\|u^{-(\alpha+\gamma_s)}(\mathcal{F}K(u)-1)\|_\infty$ is finite using the a Taylor expansion of $\mathcal{F}K$ around 0 with $(\mathcal{F}K)^{(l)}=0$ for $l=1,\dots,\lfloor\alpha+\gamma_s\rfloor$ by the order of the kernel (3.22).

For the smooth part of ζ Plancherel's identity yields

$$\begin{aligned}&\int|\zeta^c(x)(K_h*f-f)(x)|\mathrm{d}x\\ &=\frac{1}{2\pi}\int|\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](\mathrm{Id}+\mathrm{D})\{(\mathcal{F}K(hu)-1)\,\mathcal{F}f(-u)\}|\mathrm{d}u\\ &\leqslant\int|\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](\mathcal{F}K(hu)-1+h\,\mathcal{F}[ixK](hu))\,\mathcal{F}f(-u)|\mathrm{d}u\\ &\quad+\int|\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](\mathcal{F}K(hu)-1)\,\mathcal{F}[ixf](-u)|\mathrm{d}u.\end{aligned}$$

The first term can be estimated as before and for the second term we note that $xf(x)\in L^2(\mathbb{R})=H^0(\mathbb{R})$ by Assumption 3.C(i) such that the additional smoothness of $\frac{1}{ix+1}\zeta^c(x)$ yields the right order. Therefore, the second term in (3.50) is up to a constant smaller than $h^{\alpha+\gamma_s}$ and thus by the choice of h, the bias term is of order $o(n^{-1/2})$.

Now, we consider the stochastic error term in (3.50). Using $\zeta^s\in L^2$ and $\zeta^c\in L^\infty$, we can apply the smoothed adjoint equality (3.48) and obtain for the stochastic error term

$$\int \zeta(x)\,\mathcal{F}^{-1}\Big[\mathcal{F}\,K_h\frac{\varphi_n-\varphi}{\varphi_\varepsilon}\Big](x)\mathrm{d}x = \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,K_h]*\zeta(x)(P_{\nu,n}-P_\nu)(\mathrm{d}x) \tag{3.51}$$

with the empirical measure $P_{\nu,n}=\frac{1}{n}\sum_{j=1}^n\delta_{Y_j}$. Therefore, it suffices for the convergence of the finite dimensional distributions to bound the term

$$\sup_{h\in(0,1)}\int\Big|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,K_h]*\zeta(x)\Big|^{2+\delta}P_\nu(\mathrm{d}x), \tag{3.52}$$

for any function $\zeta\in Z^{\gamma_s,\gamma_c}$. Then the stochastic error term converges in distribution to a normal random variable by the central limit theorem under the Lyapunov condition (i.e., Klenke, 2007, Thm. 15.43 together with Lem. 15.41).

First, note that the moment conditions in Assumptions 3.C and 3.D and the estimate

$$|x|^pf_Y(x)\leqslant\int|x-y+y|^pf(x-y)f_\varepsilon(y)\mathrm{d}y\lesssim(|y|^pf)*f_\varepsilon+f*(|y|^pf_\varepsilon),$$

for $x\in\mathbb{R}$, $p\geqslant1$, yield finite $(2+\delta)$th moments for P_ν since

$$\int|x|^{2+\delta}f_Y(x)\mathrm{d}x\lesssim\||x|^{2+\delta}f\|_{L^1}\|f_\varepsilon\|_{L^1}+\|f\|_{L^1}\||x|^{2+\delta}f_\varepsilon\|_{L^1}<\infty. \tag{3.53}$$

To bound (3.52), we estimate separately all three terms in the decomposition (3.25) considering $\zeta*K_h$ instead of ζ. The continuity and linearity of the Fourier multiplier $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]$, which was shown in Lemma 3.41(i), yield for the first term in (3.25)

$$\begin{aligned}\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{H^\delta}&=\|\,\mathcal{F}^{-1}\,[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}[\zeta^s*K_h]]\,\|_{B^\delta_{2,2}}\\&\lesssim\|\zeta^s*K_h\|_{B^{\beta+\delta}_{2,2}}\lesssim\|\zeta^s\|_{H^{\beta+\delta}},\end{aligned}$$

where the last inequality holds by $\|\,\mathcal{F}\,K_h\|_\infty\leqslant\|K\|_{L^1}$. Using the boundedness of f_Y and the continuous Sobolev embedding $H^{\delta/4}(\mathbb{R})\subseteq L^{2+\delta}(\mathbb{R})$ by (A.4), (A.7) and (A.6), we obtain

$$\begin{aligned}&\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{L^{2+\delta}(P_\nu)}\\&\quad\lesssim\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{L^{2+\delta}}\\&\quad\lesssim\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{H^\delta}\lesssim\|\zeta^s\|_{H^{\beta+\delta}}.\end{aligned} \tag{3.54}$$

To estimate the second term in (3.25), we use the Cauchy–Schwarz inequality and Assumption 3.D(iii):

$$\begin{aligned}&\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](u)\,\mathcal{F}\,K_h(u)]\|_\infty\\&\quad\leqslant\|\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c]\,\mathcal{F}\,K_h(u)\|_{L^1}\\&\quad\lesssim\|(1+|u|)^{-1/2-\beta-\delta}\varphi_\varepsilon^{-1}(-u)\|_{L^2}\|(1+|u|)^{1/2+\beta+\delta}\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)]\|_{L^2}\\&\quad\lesssim\|\tfrac{1}{ix+1}\zeta^c(x)\|_{H^{1/2+\beta+\delta}}.\end{aligned}$$

Thus $\int(1+x^2)^{(2+\delta)/2} f_Y(x)\mathrm{d}x<\infty$ from (3.53) yields

$$\|(1+ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)\,\mathcal{F}\,K_h(u)](x)\|_{L^{2+\delta}(P_\nu)}\lesssim\|\tfrac{1}{ix+1}\zeta^c(x)\|_{H^{1/2+\beta+\delta}}. \tag{3.55}$$

The last term in the decomposition (3.25) can be estimated similarly using the Cauchy–Schwarz inequality and Assumption 3.D(iii) for $(\varphi^{-1})'$

$$\begin{aligned}
&\|\,\mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](u)\,\mathcal{F}\,K_h(u)]\|_{L^{2+\delta}(P_\nu)}\\
&\quad\lesssim\|(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](u)\|_{L^1}\\
&\quad\leqslant\|(1+|u|)^{1/2-\beta-\delta}(\varphi_\varepsilon^{-1})'\|_{L^2}\|(1+|u|)^{-1/2+\beta+\delta}\,\mathcal{F}^{-1}[\tfrac{1}{ix+1}\zeta^c(x)](u)\|_{L^2}\\
&\quad\lesssim\|\tfrac{1}{ix+1}\zeta^c(x)\|_{H^{-1/2+\beta+\delta}}.
\end{aligned} \tag{3.56}$$

Combining (3.54), (3.55) and (3.56), we obtain

$$\sup_{h\in(0,1)}\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,K_h]*\zeta(x)\|_{L^{2+\delta}(P_\nu)}\lesssim\|\zeta\|_{Z^{\beta+\delta,1/2+\beta+\delta}}, \tag{3.57}$$

which is finite for δ small enough satisfying $\beta+\delta\leqslant\gamma_s$ and $1/2+\beta+\delta\leqslant\gamma_c$. Since $\mathcal{F}\,K_h$ converges pointwise to one and $|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,K_h]*\zeta(x)|^2$ is uniformly integrable by the bound of the $2+\delta$ moments, the variance converges to

$$\int\Big|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]*\zeta(x)\Big|^2\,P_\nu(\mathrm{d}x).$$

Step 2: We show asymptotic linearity of the estimator $\widehat{\vartheta}=\langle\zeta,\widehat{f}_h\rangle$ with influence function $(x,P_\nu)\mapsto\int(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]*\zeta)(\mathrm{d}\delta_x-\mathrm{d}P_\nu)$. Regularity of the estimator follows then from the regularity of the model (P_{ν_t}) and Le Cam's third lemma (cf. van der Vaart, 1998, Lem. 25.23).

The estimate of the bias of $\widehat{\vartheta}_h$ in Step 1 yields

$$\begin{aligned}
\widehat{\vartheta}_h=&\vartheta+\int\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,K_h]*\zeta(y)(P_{\nu,n}-P_\nu)(\mathrm{d}y)+o_P(n^{-1/2})\\
=&\vartheta+\int\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]*\zeta(y)(P_{\nu,n}-P_\nu)(\mathrm{d}y)\\
&+\int\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)(\mathcal{F}\,K_h-1)]*\zeta(y)(P_{\nu,n}-P_\nu)(\mathrm{d}y)+o_P(n^{-1/2}).
\end{aligned}$$

Since

$$\mathbb{E}\left[\left|\int(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]*\zeta)(\mathrm{d}\delta_x-\mathrm{d}P_\nu)\right|^2\right]\leqslant4\,\mathbb{E}\left[\int|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]*\zeta|^2\mathrm{d}P_\nu\right]$$

is finite and $\mathbb{E}[\int\mathcal{F}^{-1}[(\varphi_\varepsilon^{-1}(-\bullet)]*\zeta)(\mathrm{d}\delta_x-\mathrm{d}P_\nu)]=0$ by (3.57) it suffices to show

$$\int\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)(\mathcal{F}\,K_h-1)]*\zeta(y)(P_{\nu,n}-P_\nu)(\mathrm{d}y)=o_P(n^{-1/2}). \tag{3.58}$$

For convenience we write $\psi_h := \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)(\mathcal{F} K_h - 1)] * \zeta$ and let $\kappa > 0$. Since (Y_j) are independent and identically distributed, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(|n^{1/2}\int \psi_h(y)(P_{\nu,n} - P_\nu)(\mathrm{d}y)| > \kappa\Big) \leqslant \tau^{-2} n\,\mathbb{E}\Big[|\int \psi_h(y)(P_{\nu,n} - P_\nu)(\mathrm{d}y)|^2\Big] \\
&= \kappa^{-2} n\,\mathbb{E}\Big[\int\int \psi_h(y)\overline{\psi_h}(z)(P_{\nu,n} - P_\nu)(\mathrm{d}y)(P_{\nu,n} - P_\nu)(\mathrm{d}z)\Big] \\
&= \kappa^{-2} n^{-1} \sum_{j,k=1}^{n} \mathbb{E}\Big[\int\int \psi_h(y)\overline{\psi_h}(z)(\delta_{Y_j} - \mathbb{P})(\mathrm{d}y)(\delta_{Y_k} - P_\nu)(\mathrm{d}z)\Big] \\
&= \kappa^{-1}\,\mathbb{E}\Big[|\int \psi_h(y)(\delta_{Y_j} - P_\nu)(\mathrm{d}y)|^2\Big] \\
&\leqslant 4\kappa^{-1} \int |\psi_h(y)|^2 P_\nu(\mathrm{d}y).
\end{aligned}$$

By uniform integrability of ψ_h^2 with respect to P_ν by (3.57) and pointwise convergence $\psi_h \to 0$ as $h \to 0$ we conclude $\int |\psi_h(y)|^2 P_\nu(\mathrm{d}y) \to 0$ and thus (3.58). □

3.4.2. Proof of Proposition 3.31

Without loss of generality we assume $\Delta = 1$ in this and the following subsections.

For any $b \in \dot{\Theta}_\nu$ let $P_\nu*(b\nu) = (P_\nu*(b\nu))^a + (P_\nu*(b\nu))^\perp$ be the Lebesgue decomposition of $P_\nu * (b\nu)$ with respect to P_ν, that is the first and second measure are absolutely continuous and singular with respect to P_ν, respectively. $\frac{\mathrm{d}P_\nu*(b\nu)}{\mathrm{d}P_\nu}$ is then defined as the Radon–Nikodym density of $(P_\nu * (b\nu))^a$ with respect to P_ν. Therefore, $A_\nu b = \frac{\mathrm{d}P_\nu*(b\nu)}{\mathrm{d}P_\nu} - \int b\mathrm{d}\nu$ is well defined without further assumptions.

In a first step we will show L^2-differentiability of the submodels corresponding to some direction $b \in \dot{\Theta}_\vartheta \cap L^\infty(\nu)$. The extension to the whole tangent set is proved in the second step. In the last step we will see that even $P_\nu * (b\nu) \ll P_\nu$ holds true.

Step 1: Let $b \in L^1(\nu) \cap L^\infty(\nu)$. We will show that the associated model $[0,1) \ni t \mapsto P_{\nu_t}$ is L^2-differentiable at 0 with derivative $A_\nu b \in L_0^2(P_\nu)$ as defined in (3.38), that is

$$\int \Big(\frac{\mathrm{d}P_{\nu_t} - \mathrm{d}P_\nu}{t\mathrm{d}P_\nu} - A_\nu b\Big)^2 \mathrm{d}P_\nu \to 0 \quad \text{as} \quad t \to 0. \tag{3.59}$$

Note that $\frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu} \in L^2(P_\nu)$ by (3.39), which follows by a similar argument as the one following (3.60). Applying Proposition 1.199 by Witting (1985), the L^2-regularity (3.59) implies the proposed Hellinger differentiability for all $b \in \dot{\Theta}_\nu \cap L^\infty(\nu)$.

Defining the measure ν_t^c via the density $\frac{\mathrm{d}\nu_t^c}{\mathrm{d}\nu} = (k(tb) - 1) =: f_{\nu_t^c}$, we write as a consequence of (3.37) and Remark 27.3 by Sato (1999)

$$P_{\nu_t} = e^{-\nu_t^c(\mathbb{R})} \sum_{k=0}^{\infty} \frac{1}{k!}(\nu_t^c)^{*k} * P_\nu.$$

Arranging terms

$$
\begin{aligned}
&\int \Big(\frac{\mathrm{d}P_{\nu_t}-\mathrm{d}P_\nu}{t\mathrm{d}P_\nu}-A_\nu b\Big)^2\mathrm{d}P_\nu\\
=&\int \Big(\frac{\mathrm{d}P_{\nu_t}-\mathrm{d}P_\nu-t\mathrm{d}(P_\nu*(b\nu))+t(\int b\nu)\mathrm{d}P_\nu}{t\mathrm{d}P_\nu}\Big)^2\mathrm{d}P_\nu\\
=&\int \Big(\frac{(e^{-\nu_t^c(\mathbb{R})}-1+t\int b\mathrm{d}\nu)\mathrm{d}P_\nu+\mathrm{d}\big(((e^{-\nu_t^c(\mathbb{R})}f_{\nu_t^c}-tb)\nu)*P_\nu\big)}{t\mathrm{d}P_\nu}\\
&\qquad+\frac{e^{-\nu_t^c(\mathbb{R})}\sum_{k=2}^\infty(k!)^{-1}\mathrm{d}\big((\nu_t^c)^{*k}*P_\nu\big)}{t\mathrm{d}P_\nu}\Big)^2\mathrm{d}P_\nu,
\end{aligned}
$$

all three terms in the numerator turn out to be of order t^2. Note that we can dominate $|f_{\nu_t^c}|\leqslant t|b|$ as well as $|f_{\nu_t^c}-tb|\leqslant t^2|b^2|$ by $|k(y)-1|\leqslant|y|$ and $|k(y)-1-y|\leqslant y^2$. Therefore,

$$
\begin{aligned}
\Big|e^{-\nu_t^c(\mathbb{R})}-1+t\int b\mathrm{d}\nu\Big|&=\Big|\sum_{k\geqslant 2}\frac{(-\nu_t^c(\mathbb{R}))^k}{k!}-\Big(\nu_t^c(\mathbb{R})-t\int b\mathrm{d}\nu\Big)\Big|\\
&\leqslant\sum_{k\geqslant 2}\frac{t^k\|b\|_{L^1(\nu)}^k}{k!}+t^2\|b\|_{L^2(\nu)}^2\\
&\leqslant(e^{\|b\|_{L^1(\nu)}}+\|b\|_{L^2(\nu)}^2)t^2,\\
|e^{-\nu_t^c(\mathbb{R})}f_{\nu_t^c}-tb|&=\Big|(e^{-\nu_t^c(\mathbb{R})}-1)f_{\nu_t^c}+f_{\nu_t^c}-tb\Big|\\
&\leqslant t^2\big(e^{\|b\|_{L^1(\nu)}}|b|+|b|^2\big),\\
\Big|e^{-\nu_t^c(\mathbb{R})}\sum_{k=2}^\infty\frac{(\nu_t^c)^{*k}}{k!}\Big|&\leqslant t^2\sum_{k=2}^\infty\frac{(|b|\nu)^{*k}}{k!}.
\end{aligned}
$$

Hence, we estimate

$$
\begin{aligned}
&\int \Big(\frac{\mathrm{d}P_{\nu_t}-\mathrm{d}P_\nu}{t\mathrm{d}P_\nu}-A_\nu b\Big)^2\mathrm{d}P_\nu\\
\leqslant& t^2\int\Big(\frac{(e^{\|b\|_{L^1(\nu)}}+\|b\|_{L^2(\nu)}^2)\mathrm{d}P_\nu+(e^{\|b\|_{L^1(\nu)}}|b|+|b^2|)\mathrm{d}(\nu*P_\nu)}{\mathrm{d}P_\nu}\\
&\qquad+\frac{\sum_{k=2}^\infty(k!)^{-1}\mathrm{d}\big((|b|\nu)^{*k}*P_\nu\big)}{\mathrm{d}P_\nu}\Big)^2\mathrm{d}P_\nu\\
\leqslant& t^2\big(e^{\|b\|_{L^1(\nu)}}+\|b\|_{L^2(\nu)}^2\big)^2\int\Big(\frac{\mathrm{d}\Big(\sum_{k=0}^\infty((|b|+|b|^2)\nu)^{*k}/(k!)\Big)*P_\nu}{\mathrm{d}P_\nu}\Big)^2\mathrm{d}P_\nu.
\end{aligned}
$$

Introducing an infinite divisible distribution μ without diffusion component, without drift and with finite jump measure $(|b|+|b|^2)\nu$, the previous line can be written as

$$
t^2\underbrace{\big(e^{\|b\|_{L^1(\nu)}}+\|b\|_{L^2(\nu)}^2\big)^2e^{2\|b\|_{L^1(\nu)}+2\|b\|_{L^2(\nu)}^2}}_{=:C(\|b\|_{L^1(\nu)},\|b\|_{L^2(\nu)})}\int\Big(\frac{\mathrm{d}(\mu*P_\nu)}{\mathrm{d}P_\nu}\Big)^2\mathrm{d}P_\nu. \tag{3.60}
$$

Therefore, the assertion holds true provided the Hellinger integral $H_2(\mu * P_\nu, P_\nu) = \int (\mathrm{d}(\mu * P_\nu)/\mathrm{d}P_\nu)^2 \,\mathrm{d}P_\nu$ is finite. To show this, we apply the bound of Renyi's distance R_2 for infinite divisible distributions by Liese (1987, Thm. 2.6). Using that both distributions, $\mu * P_\nu$ and P_ν, have the same drift and have finite variation, we obtain (for $\alpha = 2$)

$$\tfrac{1}{2}\log H_2(\mu * P_\nu, P_\nu) = R_2(\mu * P_\nu, P_\nu) \leqslant \tfrac{1}{2}\chi^2\big((|b| + b^2 + 1)\nu, \nu\big)$$

where the χ^2-distance of the jump measures satisfies

$$\begin{aligned}\chi^2\big((|b| + b^2 + 1)\nu, \nu\big) :=& \frac{1}{2}\int \Big(\frac{\mathrm{d}((|b| + b^2 + 1)\nu)}{\mathrm{d}\nu} - 1\Big)^2 \mathrm{d}\nu \\ =& \frac{1}{2}\int \big(|b| + b^2\big)^2 \mathrm{d}\nu \leqslant \frac{1}{2}(1 + \|b\|^2_{L^\infty(\nu)})\|b\|^2_{L^2(\nu)} < \infty.\end{aligned}$$

The combination with the bound (3.60) yields

$$\int \Big(\frac{\mathrm{d}P_{\nu_t} - \mathrm{d}P_\nu}{t\mathrm{d}P_\nu} - A_\nu b\Big)^2 \mathrm{d}P_\nu \leqslant t^2 C(\|b\|_{L^1(\nu)}, \|b\|_{L^2(\nu)}) e^{\frac{1}{2}(1+\|b\|^2_{L^\infty(\nu)})\|b\|^2_{L^2(\nu)}}.$$

As $t \to 0$ this upper bound converges to zero which shows the L^2-differentiability. We conclude $\int A_\nu b \mathrm{d}P_\nu = 0$ for all $b \in \dot{\Theta}_\nu \cap L^\infty(\nu)$.

Step 2: To show continuity of $A_\nu|_{L^1(\nu)\cap L^\infty(\nu)}$, let $\varepsilon > 0$ and $b \in L^1(\nu) \cap L^\infty(\nu)$ with $\|b\|^2_{L^2(\nu)} < \varepsilon$. By (3.59), $\frac{1}{2}A_\nu b$ is the L^2-limit of $t^{-1}(\sqrt{\mathrm{d}P_{\nu_t}} - \sqrt{\mathrm{d}P_\nu})$ and thus for t small enough

$$\|A_\nu b\|^2_{L^2(P_\nu)} \leqslant \frac{2}{t^2}\int \big(\sqrt{\mathrm{d}P_{\nu_t}} - \sqrt{\mathrm{d}P_\nu}\big)^2 + \varepsilon.$$

As above Theorem 2.6 by Liese (1987) for $\alpha = 1/2$ yields the estimate for the Hellinger distance of the infinite divisible distributions

$$\begin{aligned}\int \big(\sqrt{\mathrm{d}P_{\nu_t}} - \sqrt{\mathrm{d}P_\nu}\big)^2 &\leqslant 2\Big(1 - \exp\Big(-2\int (\sqrt{\mathrm{d}\nu_t} - \sqrt{\mathrm{d}\nu})^2\Big)\Big) \\ &= 2\Big(1 - \exp\Big(-2\int (\sqrt{k(tb)} - 1)^2 \mathrm{d}\nu\Big)\Big).\end{aligned} \tag{3.61}$$

Since $|\sqrt{k(y)} - 1| \leqslant |(\sqrt{k(y)} + 1)(\sqrt{k(y)} - 1)| = |k(y) - 1| \leqslant |y|$ and $1 - e^{-y} \leqslant |y|$ for $y \in \mathbb{R}$, the previous display can be bounded by

$$2\Big(1 - \exp\Big(-2\int (tb)^2 \mathrm{d}\nu\Big)\Big) \leqslant 4t^2\|b\|^2_{L^2(\nu)}.$$

Because $\varepsilon > 0$ was arbitrary, we conclude $\|A_\nu b\|_{L^2(P_\nu)} \lesssim \|b\|_{L^2(\nu)}$, which is equivalent to the continuity of the linear operator $A_\nu|_{L^1(\nu)\cap L^\infty(\nu)}$. Since $L^1(\nu) \cap L^\infty(\nu)$ is dense in $\dot{\Theta}_\nu$, there is a unique continuous extension A_ν on $\dot{\Theta}_\nu$ satisfying $A_\nu b = \frac{\mathrm{d}P_\nu * (b\nu) - \int b\nu \mathrm{d}P_\nu}{\mathrm{d}P_\nu}$ for all $b \in L^1(\nu) \cap L^\infty(\nu)$.

Now, for any $b \in \dot{\Theta}_\nu$ with associated path $t \mapsto \mathrm{d}\nu_t = k(tb)\mathrm{d}\nu$ and for any positive null sequence $(t_m)_{m\in\mathbb{N}}$ and let $\widetilde{b}_m \in L^1(\nu) \cap L^\infty(\nu)$ with path $t \mapsto \mathrm{d}\widetilde{\nu}_t := k(t\widetilde{b}_m)\mathrm{d}\nu$ such that $\|b-\widetilde{b}_m\|_{L^2(\nu)} \to 0$ and $\|\widetilde{b}_m\|_{L^\infty(\nu)} = o(|\log t_m|^{1/2})$ as $m\to\infty$. Then

$$\begin{aligned}
&t_m^{-2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\nu_{t_m}}}{\mathrm{d}P_\nu}}-1-\frac{t_m}{2}A_\nu b\Big)^2\mathrm{d}P_\nu\\
\leqslant&\frac{3}{t_m^2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\nu_{t_m}}}{\mathrm{d}P_\nu}}-\sqrt{\frac{\mathrm{d}P_{\widetilde{\nu}_{t_m}}}{\mathrm{d}P_\nu}}\Big)^2\mathrm{d}P_\nu+\frac{3}{t_m^2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\widetilde{\nu}_{t_m}}}{\mathrm{d}P_\nu}}-1-\frac{t_m}{2}A_\nu\widetilde{b}_m\Big)^2\mathrm{d}P_\nu\\
&+\frac{3}{4}\int\Big(A_\nu\widetilde{b}_m-A_\nu b\Big)^2\mathrm{d}P_\nu
\end{aligned}\tag{3.62}$$

The first term is the Hellinger distance between P_{ν_t} and $P_{\widetilde{\nu}_t}$, which can be bounded as in (3.61)

$$\begin{aligned}
&t_m^{-2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\nu_{t_m}}}{\mathrm{d}P_\nu}}-\sqrt{\frac{\mathrm{d}P_{\widetilde{\nu}_{t_m}}}{\mathrm{d}P_\nu}}\Big)^2\mathrm{d}P_\nu\\
&\quad\leqslant 2t_m^{-2}\Big(1-\exp\Big(-2\int\big(\sqrt{\mathrm{d}\nu_{t_m}}-\sqrt{\mathrm{d}\widetilde{\nu}_{t_m}}\big)^2\Big)\Big)\\
&\quad= 2t_m^{-2}\Big(1-\exp\Big(-2\int\Big(\sqrt{k(t_m b(x))}-\sqrt{k(t_m\widetilde{b}_m(x))}\Big)^2\nu(\mathrm{d}x)\Big)\Big).
\end{aligned}$$

An easy calculation shows $|(\sqrt{k})'(x)|\leqslant 1$ for all $x\in\mathbb{R}$ and thus the above display can be bounded by the mean value theorem

$$2t_m^{-2}\Big(1-\exp\big(-2t_m^2\|b-\widetilde{b}_m\|^2_{L^2(\nu)}\big)\Big)\leqslant 4\|b-\widetilde{b}_m\|^2_{L^2(\nu)}\to 0.$$

The second term in (3.62) converges to zero according to Step 1 provided $\|\widetilde{b}\|_{L^\infty(\nu)} = o(|\log t|^{1/2})$. Applying continuity of A_ν, the third term in (3.62) vanishes as well. Therefore, we have shown that $A_\nu b$ is the Hellinger derivative of P_{ν_t} for all $b\in\dot{\Theta}_\nu$.

Step 3: Finally, we will show $P_\nu*(b\nu)\ll P_\nu$ for all $b\in\dot{\Theta}_\nu\cap L^\infty(\nu)$. By construction $|b|\in\dot{\Theta}_\nu\cap L^\infty(\nu)$, too. Let $P_\nu*(|b|\nu)=(P_\nu*(|b|\nu))^a+(P_\nu*(|b|\nu))^\perp$ be Lebesgue's decomposition with respect to P_ν where both measures can be chosen to be nonnegative and finite. According to Step 1, $\int A_\nu|b|\mathrm{d}P_\nu=0$ which yields together with the nonnegativity of the measures and Fubini's theorem

$$\int|b|\mathrm{d}\nu=\int\frac{\mathrm{d}(P_\nu*(|b|\nu))}{\mathrm{d}P_\nu}\mathrm{d}P_\nu=\int\mathrm{d}(P_\nu*(|b|\nu))^a\leqslant\int\mathrm{d}P_\nu*(|b|\nu)=\int|b|\mathrm{d}\nu.$$

We conclude $(P_\nu*(|b|\nu))^a=P_\nu*(|b|\nu)$ or equivalently $P_\nu*(|b|\nu)\ll P_\nu$. Now for any event $A\in\mathscr{B}(\mathbb{R})$ with $P_\nu*(|b|\nu)(A)=0$ we have

$$\begin{aligned}
|P_\nu*(b\nu)(A)|&=\Big|\int_{\mathbb{R}^2}\mathbb{1}_A(x+y)b(x)\nu(\mathrm{d}x)P_\nu(\mathrm{d}y)\Big|\\
&\leqslant\int_{\mathbb{R}^2}\mathbb{1}_A(x+y)|b(x)|\nu(\mathrm{d}x)P_\nu(\mathrm{d}y)=P_\nu*(|b|\nu)(A)=0.
\end{aligned}$$

Consequently, $P_\nu*(b\nu)\ll P_\nu*(|b|\nu)\ll P_\nu$. □

3.4.3. Proof of Lemma 3.35

(i) We will determine the adjoint score operator and its inverse on the subsets $\mathcal{G}$ and $\mathcal{H}$ as defined in (3.43). In the case of infinite jump activity the application of Fubini's theorem in (3.42) holds as well. Hence, the adjoint score operator on $\mathcal{G}$ is given by $A_\nu^\star g = P_\nu(-\bullet) * g$. To verify that $A_\nu^\star|_{\mathcal{G}}$ is well defined, we note first that by the Sobolev embedding any $g \in \mathcal{G}$ has a version in $C^1(\mathbb{R})$. Throughout we can identify g with this smooth version. Then, we obtain $A_\nu^\star g(0) = \int g \mathrm{d}P_\nu = 0$ and $\|(A_\nu^\star g)^{(l)}\|_\infty = \|P_\nu(-\bullet) * (g^{(l)})\|_\infty \leqslant \|g^{(l)}\|_\infty \leqslant \|g\|_{C^1}$ for $l = 0, 1$. Hence, $A_\nu^\star g$ is a bounded function and

$$\begin{aligned}\int |A_\nu^\star g(x)|\nu(x)\mathrm{d}x &\leqslant \int (\|A_\nu^\star g\|_\infty \wedge (|x|\|(A_\nu^\star g)'\|_\infty))\mathrm{d}\nu(x)\\ &\leqslant \|g\|_{C^1}\int (1 \wedge |x|)\mathrm{d}\nu(x).\end{aligned}$$

A similar estimate holds for $L^2(\nu)$. Therefore, $A_\nu^\star g \in L^1(\nu) \cap L^\infty(\mathbb{R}) \subseteq \dot{\Theta}_\nu$. Owing to $\|\varphi_\nu\|_\infty \leqslant 1$, it holds for any $s > 0$

$$\big\|(1 + |u|^2)^{s/2}\,\mathcal{F}[A_\nu^\star g](u)\big\|_{L^2} = \big\|(1 + |u|^2)^{s/2}\varphi_\nu(-u)\,\mathcal{F}\,g(u)\big\|_{L^2} \leqslant \|g\|_{H^s} < \infty.$$

We conclude $\operatorname{ran} A_\nu^\star|_{\mathcal{G}} \subseteq \mathcal{H}$.

Let us show now that the inverse adjoint score operator as given in (3.45) is well defined on $\mathcal{H}$. Applying the assumption $|\varphi_\nu(u)| \gtrsim (1 + |u|)^{-\beta}$, we obtain for all $b \in \mathcal{H}$ and $s > 0$

$$\begin{aligned}\big\|(1 + |u|^2)^{s/2}\,\mathcal{F}[(A_\nu^\star)^{-1}b](u)\big\|_{L^2} &= \big\|(1 + |u|^2)^{s/2}\,\mathcal{F}\,b(u)/\varphi_\nu(-u)\big\|_{L^2}\\ &\lesssim \big\|(1 + |u|^2)^{(s+\beta)/2}\,\mathcal{F}\,b(u)\big\|_{L^2}\\ &\leqslant \|b\|_{H^{s+\beta}} < \infty.\end{aligned}$$

Therefore, $(A_\nu^\star)^{-1}b \in H^\infty(\mathbb{R})$ and the Sobolev embedding yields $\|(A_\nu^\star)^{-1}b\|_{L^2(P_\nu)} \leqslant \|(A_\nu^\star)^{-1}b\|_\infty \leqslant \|(A_\nu^\star)^{-1}b\|_{C^s} < \infty$. It remains to verify the condition $\int (A_\nu^\star)^{-1}b\mathrm{d}P_\nu = 0$. By construction

$$\begin{aligned}\int (A_\nu^\star)^{-1}b\mathrm{d}P_\nu &= (((A_\nu^\star)^{-1}b) * P_\nu)(0)\\ &= \big(\mathcal{F}^{-1}\left[\mathcal{F}\,b/\varphi_\nu(-\bullet)\right] * P_\nu\big)(0) = b(0),\end{aligned}$$

where the last equality is clear in distributional sense and can be shown via integration against test functions. Since $b(0) = 0$ for all $b \in \mathcal{H}$, we conclude $\operatorname{ran}(A_\nu^\star)^{-1}|_{\mathcal{H}} \subseteq \mathcal{G}$.

By construction $g = (A_\nu^\star)^{-1}A_\nu^\star g$ and $b = A_\nu^\star(A_\nu^\star)^{-1}b$ for all $g \in \mathcal{G}, b \in \mathcal{H}$ which proves that $A_\nu^\star|_{\mathcal{G}}$ is a bijection from $\mathcal{G}$ onto $\mathcal{H}$.

(ii) Let us show that $\mathcal{G}$ is dense in $L_0^2(P_\nu)$. Since the Borel σ-field is generated by $\mathcal{E} := \{[a, b] : -\infty < a < b < \infty\}$, the set of indicator functions $\{\mathbb{1}_E : E \in \mathcal{E}\}$ is dense in $L^2(P_\nu)$. Hence, it suffices to approximate in $L^2(P_\nu)$-sense the indicators $\mathbb{1}_E, E \in \mathcal{E}$, by smooth $L^2(\mathbb{R})$-integrable functions. Let $\varepsilon > 0$ be arbitrary and let us denote the distribution function of P_ν by $F(x) := P_\nu((-\infty, x]), x \in \mathbb{R}$. Since F is right continuous with left limits, for all $a < b$ there are $a' < a, b < b'$ such that

$$P_\nu\big((a', b'] \setminus [a, b]\big) = P_\nu\big((a', a)\big) + P_\nu\big((b, b']\big) = F(a-) - F(a') + F(b') - F(b) < \varepsilon.$$

Therefore, for any $A \in \mathcal{E}$ there is a bounded set $B \in \mathscr{B}(\mathbb{R})$ satisfying $A \subseteq B$, $P_\nu(B \setminus A) < \varepsilon$ and the distance between $x \in A$ and $\mathbb{R} \setminus B$ is strictly positive. Consequently, there is some non-negative $\psi \in C^\infty(\mathbb{R})$ with $\psi(x) = 1$ for $x \in A$, $\|\psi\|_\infty < 1$ and $\operatorname{supp} \psi \subseteq B$. Obviously, ψ is contained in $H^\infty(\mathbb{R}) \cap L^2(P_\nu)$ and $\|\mathbb{1}_A - \psi\|_{L^2(P_\nu)} < \sqrt{\varepsilon}$. Since $L^2_0(P_\nu) = (\operatorname{lin} 1)^\perp$ is a closed subspace of $L^2(P_\nu)$, we conclude that $\mathcal{G}$ is dense in $L^2_0(P_\nu)$.

Continuity of $A^\star_\nu$ follows from the continuity of A_ν, which was shown in Proposition 3.31. Hence, $A^\star_\nu$ is uniquely given by the continuous extension of $A^\star_\nu|_{\mathcal{G}}$ to $L^2_0(P_\nu)$. □

3.4.4. Proof of Proposition 3.39

Taking the derivative of the Lévy-Khintchine formula (3.32), we obtain

$$\varphi'_\nu(u) = \varphi_\nu(u)\big(i\gamma_0 + \mathcal{F}[x\nu](u)\big), \quad u \in \mathbb{R}\,.$$

In a first step we will show that the drift can be discarded, which was also the case for the upper bound by Nickl and Reiß (2012). Since Lemma 3.35 shows that the inverse adjoint score operator is given by $\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)]$ on the smooth subset $\mathcal{G}$, we study the mapping properties of this deconvolution operator in Step 2. Finally in Step 3, we apply the characterization in Proposition 3.12 to prove that $Z^\beta(\mathbb{R}) \subseteq \operatorname{ran} A^\star_\nu$ and to determine $(A^\star_\nu)^{-1}$ on $Z^\beta(\mathbb{R})$.

Step 1: Let us show that $\gamma_0 = 0$ can be assumed, meaning that the process L has no drift. For any $\gamma_0 \in \mathbb{R}$ consider the infinitely divisible distribution $\widetilde{P}_\nu := P_\nu * \delta_{-\gamma_0}$. Then the following map is an isomorphism

$$\Phi: \quad L^2_0(P_\nu) \to L^2_0(\widetilde{P}_\nu), \quad g \mapsto g(\bullet + \gamma_0).$$

Lemma 3.35 determines the adjoint score operator $\widetilde{A}^\star_\nu$ which corresponds to $\widetilde{P}_\nu$. Also by Lemma 3.35 we see for $g \in \mathcal{G}$ that $A^\star_\nu g = P_\nu(-\bullet) * g = \widetilde{P}_\nu(-\bullet) * g(\bullet + \gamma_0)$. Therefore, $A^\star_\nu = \widetilde{A}^\star_\nu \circ \Phi$ which implies

$$\operatorname{ran} A^\star_\nu = \operatorname{ran}\big(\widetilde{A}^\star_\nu \circ \Phi\big) = \operatorname{ran} \widetilde{A}^\star_\nu.$$

Hence, for the rest of the proof suppose $\gamma_0 = 0$.

Step 2: The aim of this step is to show for $\zeta = \zeta^s + \zeta^c$ and any $\beta^+ > \beta$

$$\|\,\mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,\zeta]\|_{L^2(P_\nu)} \lesssim \|\zeta^s\|_{H^{\beta+}} + \big\|x^{-1}\zeta^s(x)\big\|_{H^{\beta+}} + \|\zeta^c\|_{C^{\beta+}}. \tag{3.63}$$

To this end note that Assumption 3.E together with Lemma 5.2 yields for any $\beta^+ > \beta$, due to $\gamma_0 = 0$,

$$|\varphi_\nu^{-1}(u)| \lesssim (1+|u|)^{\beta^+} \quad \text{and} \quad |(\varphi_\nu^{-1})'(u)| \lesssim (1+|u|)^{\beta^+-1}$$

and Theorem 5.5 shows that for all $s \in \mathbb{R}, p, q \in [1,\infty]$ the linear map

$$B^{s+\beta^+/2}_{p,q}(\mathbb{R}) \to B^s_{p,q}(\mathbb{R}), \quad f \mapsto \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,f] \tag{3.64}$$

is bounded. This yields for any $\varepsilon > 0$ and $\zeta^c \in C^{\beta^+}(\mathbb{R})$

$$\|F^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,\zeta^c]\|_{L^2(P_\nu)} \lesssim \|\,\mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,\zeta^c]\|_\infty \lesssim \|\zeta^c\|_{B^{\beta+/2}_{\infty,1}} \lesssim \|\zeta^c\|_{C^{\beta+}}. \tag{3.65}$$

For the singular part we apply a similar decomposition as (3.25). Integration by parts yields

$$\begin{aligned}\mathcal{F}^{-1}\Big[\frac{\mathcal{F}\,\zeta^s}{\varphi_\nu(-\bullet)}\Big] &= \mathcal{F}^{-1}\Big[\frac{(\mathcal{F}[\frac{1}{ix}\zeta^s(x)])'}{\varphi_\nu(-\bullet)}\Big]\\ &= ix\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big] + \mathcal{F}^{-1}\Big[\mathcal{F}\left[\tfrac{1}{ix}\zeta^s(x)\right](\varphi_\nu^{-1})'(-\bullet)\Big].\end{aligned}$$

Note that $1/\varphi_\nu$ is a Fourier multiplier from $H^{\beta+}$ into $H^0 = L^2(\mathbb{R})$ on the assumption $|\varphi_\nu(u)| \gtrsim (1+|u|)^{-\beta+}$. Similarly, $(\varphi_\nu^{-1})'$ is a Fourier multiplier from $H^{\beta+}$ into H^1. Hence, for $\frac{1}{i\bullet}\zeta^s \in H^{\beta+}$

$$\begin{aligned}\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big] &\in L^2(\mathbb{R}),\\ \mathcal{F}^{-1}\Big[\mathcal{F}\left[\tfrac{1}{ix}\zeta^s(x)\right](\varphi_\nu^{-1})'(-\bullet)\Big] &\in H^1(\mathbb{R}) \subseteq C^0(\mathbb{R}),\end{aligned}$$

where the last inclusion holds by the Sobolev embedding. Moreover,

$$ix\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big] = \mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big] - \mathcal{F}^{-1}\Big[\mathcal{F}\left[\tfrac{1}{ix}\zeta^s(x)\right](\varphi_\nu^{-1})'(-\bullet)\Big]$$

which is an $L^2(\mathbb{R})$-function. Applying Lemma 4(a) by Nickl and Reiß (2012), the distribution $xP_\nu(\mathrm{d}x)$ has a bounded Lebesgue density and which yields together with $|x|^2 \lesssim |x||1+ix|^2$ and the continuous embeddings above

$$\begin{aligned}\|\,\mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s]\|_{L^2(P_\nu)} \leqslant &\int\Big|ix\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big](x)\Big|^2\mathrm{d}P_\nu(x)\\ &+\int\Big|\,\mathcal{F}^{-1}\Big[\mathcal{F}\left[\tfrac{1}{ix}\zeta^s(x)\right](\varphi_\nu^{-1})'(-\bullet)\Big](x)\Big|^2\mathrm{d}P_\nu(x)\\ \lesssim&\Big\|(1+ix)\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big](x)\Big\|_{L^2}\\ &+\Big\|\,\mathcal{F}^{-1}\Big[\mathcal{F}\left[\tfrac{1}{ix}\zeta^s(x)\right](\varphi_\nu^{-1})'(-\bullet)\Big](x)\Big\|_\infty\\ \lesssim&\big\|\tfrac{1}{x}\zeta^s(x)\big\|_{H^{\beta+}} + \|\zeta^s\|_{H^{\beta+}}.\end{aligned}$$

Combining with (3.65), we get (3.63).

Step 3: Define the sets

$$\mathcal{G}' := C^\infty(\mathbb{R}) \cap L^2_0(P_\nu) \quad \text{and} \quad \mathcal{H}' := \{b \in C^\infty(\mathbb{R}) | b(0) = 0\} \cap \dot{\Theta}_\nu,$$

which are larger than $\mathcal{G}$ an $\mathcal{H}$ from above. Using the Fourier multiplier property on Besov spaces (3.64) and an analogous result for the Fourier multiplier $\varphi_\nu(-\bullet)$, we obtain

$$\|P_\nu(-\bullet) * f\|_{C^{s'}} \lesssim \|f\|_{C^{s'}} \quad \text{and} \quad \|F^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,f]\|_{C^s} \lesssim \|f\|_{C^{s'}}$$

for any $s > 0$ and $f \in C^{s'}$ for $s' > s + \beta$. Therefore, following the lines of the proof of Lemma 3.35(i), we see that $A_\nu^\star|_{\mathcal{G}'}$ is given by $A_\nu^\star g = P_\nu(-\bullet) * g$ for $g \in \mathcal{G}'$ and that $A_\nu^\star|_{\mathcal{G}'}$ is a bijection from $\mathcal{G}'$ onto $\mathcal{H}'$ with inverse $(A_\nu^\star|_{\mathcal{G}'})^{-1}b = F^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,b]$ for $b \in \mathcal{H}'$.

By Proposition 3.12 for any ζ a necessary and sufficient condition to be in the range of $A_\nu^\star$ is the existence of a sequence $(\chi_m)_{m\in\mathbb{N}} \subseteq \mathcal{H}'$ such that $\chi_m \to \zeta$ in $L^2(\nu)$ and $(A_\nu^\star)^{-1}\chi_m$ converges in $L^2(P_\nu)$. Now, for any $\zeta \in Z^{\beta^+} \cap L^1(\nu) \cap L^2(\nu)$ we find $\chi_m = \chi_m^s + \chi_m^c$ with $\chi_m^s, \chi_m^c \in \mathcal{H}'$ satisfying $\chi_m^s \to \zeta^s$ and $\chi_m^c \to \zeta^c$ in $L^2(\nu)$ as well as

$$\Big\| \mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\zeta - \chi_m]}{\varphi_\nu(-\bullet)}\Big]\Big\|_{L^2(P_\nu)} \lesssim \|\zeta^s - \chi_m^s\|_{H^{\beta+}} + \|x^{-1}(\zeta^s - \chi_m^s)(x)\|_{H^{\beta+}} + \|\zeta^c - \chi_m^c\|_{C^{\beta+}} \to 0$$

for $m \to \infty$ owing to (3.63). Hence, $\zeta \in \operatorname{ran} A_\nu^\star$ with $(A_\nu^\star)^{-1}\zeta = \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\zeta] = \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)] * \zeta$. □

4. Adaptive quantile estimation in deconvolution

Let us shortly recall the deconvolution model. Let $X_1, \dots, X_n$ be independent random variables with a common Lebesgue density $f : \mathbb{R} \to \mathbb{R}$. We observe the random variables

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n, \tag{4.1}$$

with i.i.d. error variables ε_j, independent of (X_j) and with the Lebesgue density f_ε. For $\tau \in (0,1)$ the objective is to estimate the τ-quantile q_τ of the population X from the observations $Y_1, \dots, Y_n$. Additionally, we observe a sample from f_ε given by

$$\varepsilon_1^*, \dots, \varepsilon_m^*, \qquad m \in \mathbb{N}\,. \tag{4.2}$$

We will not assume that the observations (ε_k^*) are independent from (Y_j).

Assuming that the distribution of the measurement error is completely known, Carroll and Hall (1988) have constructed a kernel density estimator based on the empirical characteristic function $\varphi_n(u) := \frac{1}{n}\sum_{j=1}^n e^{iuY_j}, u \in \mathbb{R}$. Neumann (1997) has proposed the following density estimator of f for the case of unknown error distributions:

$$\widetilde{f}_h(x) := \mathcal{F}^{-1}\Big[\frac{\varphi_n(u)\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\mathbb{1}_{\{|\varphi_{\varepsilon,m}(u)|\geq m^{-1/2}\}}\Big](x), \quad x \in \mathbb{R}, \tag{4.3}$$

where φ_K is the Fourier transform of a kernel K, $h > 0$ is its bandwidth and the characteristic function of the error distribution φ_ε is estimated by its empirical counterpart $\varphi_{\varepsilon,m}(u) := \frac{1}{m}\sum_{k=1}^m e^{iu\varepsilon_k^*}, u \in \mathbb{R}$. Obviously, $\widetilde{f}_h$ depends on the sample sizes n and m which is suppressed in the notation. Applying the plug-in approach, we define the distribution function estimator

$$\widetilde{F}_h(\eta) := \int_{-\infty}^{\eta} \widetilde{f}_h(x)\mathrm{d}x, \quad \eta \in \mathbb{R}\,.$$

Since Fan (1991b) required a truncation of the integral and Hall and Lahiri (2008) as well as Dattner et al. (2011) used a different distribution function estimator, it was left open whether the plug-in approach can be applied for this estimation problem. One of the main contributions in this chapter is to show that $\widetilde{F}_h$ indeed achieves the optimal rates.

Our estimator for the quantile q_τ is then given by the minimum-contrast estimator

$$\widetilde{q}_{\tau,h} := \mathrm{argmin}_{\eta\in[-U_n,U_n]}\,|\widetilde{M}_h(\eta)| \quad \text{with} \quad \widetilde{M}_h(\eta) = \int_{-\infty}^{\eta} \widetilde{f}_h(x)\mathrm{d}x - \tau \tag{4.4}$$

for some $U_n \to \infty$. As the very first step we will show that these estimators are indeed well-defined with overwhelming probability. In this chapter we pursue the analysis for error distributions whose characteristic function decays polynomially. As shown by Fan (1991b), these so-called ordinary smooth errors lead to mildly ill-posed estimation problems. They are mathematically more challenging than the so-called super-smooth errors, which we discuss briefly in Section 4.1.3.

In Section 4.1 we show the convergence rates for the estimators. Since the deconvolution operator $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$ is not observable, we have to study the estimated counterpart $\mathcal{F}^{-1}[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\mathbb{1}_{\{|\varphi_{\varepsilon,m}(u)|\geq m^{-1/2}\}}]$. As a random Fourier multiplier it preserves the mapping properties of the deterministic $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$, but its operator norm turns out to be (slightly) larger.

A lower bound result establishes that the rates under a local Hölder condition are indeed minimax optimal. Surprisingly, the dependence of the minimax rate on the error sample size m is completely different from the case of global Sobolev restrictions like in Neumann (1997). The proof enlightens this interplay between the decay of one characteristic function and estimation error in the other sample for both, the (Y_j) and the (ε_j).

The adaptive bandwidth choice is developed in Section 4.2. To this end a variant of Lepski's method is applied, but because of the unknown and possibly dependent error distribution a much more refined analysis is needed to establish that the resulting adaptive quantile estimator is (up to logarithmic factors) still rate optimal.

In Section 4.3 we implement our estimation procedure and present simulation results which show a good performance of the estimator. In a real data example we consider multiple blood pressure measurement data from different patients. Here, a measurement error is clearly present, but of unknown distribution and we have to estimate it by taking patient-wise differences. The completely data-driven method yields reasonable quantile estimates which differ from the sample quantiles of the directly measured (Y_j). All proofs are postponed to Section 4.4.

4.1. Convergence rates

4.1.1. Setting and upper bounds

Similar to Assumption 3.B, let the kernel satisfy the following

Assumption 4.A. Let the kernel function $K \in L^1(\mathbb{R})$ with Fourier transform $\varphi_K := \mathcal{F} K$ satisfy

(i) $\operatorname{supp} \varphi_K \subseteq [-1,1]$ and

(ii) K has order $\ell \in \mathbb{N}$, i.e., for $k = 1, \dots \ell$

$$\int K(x)\mathrm{d}x = 1, \quad \int x^k K(x)\mathrm{d}x = 0 \quad \text{and} \quad \int |K(x)||x|^{\ell+1}\mathrm{d}x < \infty.$$

Examples of such kernels can be obtained by taking $\mathcal{F} K$ to be a symmetric function in $C^\infty(\mathbb{R})$ which is supported in $[-1,1]$ and constant to one in a neighborhood of zero. The resulting kernels are called flat top kernels and were used in deconvolution problems, for example, by Bissantz et al. (2007). An explicit example is given in (2.15).

By construction the quantile estimator $\widetilde{q}_{\tau,h}$ is the approximated solution of the estimating equation

$$0 = \widetilde{M}_h(\eta) = \int_{-\infty}^{\eta} \widetilde{f}_h(x)\mathrm{d}x - \tau. \tag{4.5}$$

If a solution exists, it does not have to be unique since $\widetilde{f}_h$ is not necessarily non-negative. Nevertheless, any choice converges to the true quantile, assuming the latter is unique. Before, integrability of $\widetilde{f}_h$ was an open problem, which we shall settle now.

Lemma 4.1. *Grant Assumption 4.A with $\ell = 0$. On the event*

$$B_\varepsilon(h) := \Big\{ \inf_{u\in[-1/h,1/h]} |\varphi_{\varepsilon,m}(u)| \geqslant m^{-1/2}|\log h|^{3/2} \Big\} \tag{4.6}$$

we have $\widetilde{f}_h \in L^1(\mathbb{R})$ and the estimating equation (4.5) *has a solution.*

Therefore, a truncation of the integral as used by Fan (1991b) is not necessary, implying that no tail condition on f is required. Although $\|\widetilde{f}_h\|_{L^1}$ is finite, it depends on the observations as well as through h on n, m. To quantify the behavior of $\widetilde{f}_h$ more precisely, our analysis relies on the following much stronger result.

Lemma 4.2. *Grant Assumption 4.A with $\ell = 0$. For some $\beta, R > 0$ suppose $\mathbb{E}[(\varepsilon_k^*)^4] \leqslant R$ and*

$$|\varphi_\varepsilon(u)|^{-1} \leqslant R(1+|u|)^\beta \quad \text{and} \quad |\varphi_\varepsilon'(u)| \leqslant R(1+|u|)^{-\beta-1}$$

as well as $mh^{2\beta+1} \to \infty$. Then there exists a finite random variable $\mathcal{E}_h$ which is $\mathcal{O}_P\big(1 \vee \frac{1}{m^{1/2}h^{\beta+1}}\big)$ with the constant depending only on β and R, such that for any $s > \beta^+ > \beta$ on the event $B_\varepsilon(h)$ from (4.6)

$$\Big\| \mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big] * \psi \Big\|_{C^{s-\beta^+}} \leqslant \mathcal{E}_h \|\psi\|_{C^s} \quad \textit{for all} \quad \psi \in C^s(\mathbb{R}).$$

This lemma states that the random Fourier multiplication operator $C^s(\mathbb{R}) \ni \psi \mapsto \mathcal{F}^{-1}\big[\frac{\varphi_K(hu)\,\mathcal{F}\psi(u)}{\varphi_{\varepsilon,m}(u)}\big] \in C^{s-\beta^+}(\mathbb{R})$ has a norm bound $\mathcal{O}_P\big(1 \vee \frac{1}{m^{1/2}h^{\beta+1}}\big)$ on the event $B_\varepsilon(h)$. The deterministic counterpart of this lemma was proved in Söhl and Trabs (2012b). The condition on the derivative φ_ε' is natural in the context of Fourier multipliers and is usually satisfied for distributions with polynomial decaying characteristic functions, e.g. for Gamma distributions with shape parameter $\beta > 0$.

Remark 4.3. Depending only on the observations, condition (4.6) can be verified by the practitioner for a given bandwidth h. On the assumptions of Lemma 4.2 Talagrand's inequality yields $P(B_\varepsilon(h)) \geqslant 1 - 2e^{-mh^{2\beta+1}}$ (cf. Lemma 4.13 and (4.64) below). Therefore, with overwhelming probability $B_\varepsilon(h)$ holds true and the estimating equation (4.5) is rigorously defined.

Before we start with the error analysis, let us describe the class of densities we are interested in. Let $\mathcal{Q}(R)$ denote the set of all probability densities on $\mathbb{R}$ which are uniformly bounded by $R > 0$ and recall the definition of the H"older space $C^\alpha(I, R)$

from (A.3). Following the minimax paradigm, we consider for $R, r, \zeta, U > 0$ and the smoothness index $\alpha > 0$ the classes

$$\begin{aligned}\mathcal{C}^\alpha(R,r,\zeta) &:= \bigcup_{U\in\mathbb{N}} \mathcal{C}^\alpha(R,r,\zeta,U) \quad \text{and}\\ \mathcal{C}^\alpha(R,r,\zeta,U) &:= \Big\{f \in \mathcal{Q}(R)\Big| f \text{ has a } \tau\text{-quantile } q_\tau \in (-U,U) \text{ such that}\\ &\qquad f \in C^\alpha([q_\tau-\zeta, q_\tau+\zeta], R) \text{ and } f(q_\tau) \geqslant r\Big\}.\end{aligned} \tag{4.7}$$

In contrast to Dattner et al. (2011), the smoothness is measured locally in a Hölder scale and not globally by decay conditions of the Fourier transform of f. The former is more natural since both, the distribution function and the quantile function are estimated pointwise. Note that the quantile q_τ is unique, given the assumption $f(q_\tau) > 0$. Recalling that we write $\varphi_\varepsilon := \mathcal{F} f_\varepsilon$, the conditions in Lemma 4.2 motivate the definition of the class of error densities

$$\begin{aligned}\mathcal{D}^\beta(R,\gamma) := \Big\{f_\varepsilon \in \mathcal{Q}(\infty)\Big| \tfrac{1}{R}(1+|u|)^{-\beta} \leqslant |\mathcal{F} f_\varepsilon(u)| \leqslant R(1+|u|)^{-\beta},\\ |(\mathcal{F} f_\varepsilon)'(u)| \leqslant R(1+|u|)^{-1-\beta}, \|x^\gamma f_\varepsilon(x)\|_{L^1} \leqslant R\Big\}\end{aligned} \tag{4.8}$$

for some moment $\gamma \geqslant 0$ and we use the same constant R as above for convenience.

Remark 4.4. The upper and lower bounds for $|\varphi_\varepsilon(u)|$ in $\mathcal{D}^\beta(R,\gamma)$ are standard assumptions in deconvolution and are used for deriving lower bounds for the estimation problem as well as upper bounds for the risk of the estimators. Specifically, these bounds correspond to ordinary smooth error distributions (Fan, 1991b), see Section 4.1.3 below for the super-smooth case.

Applying the plug-in approach, we need to integrate the density estimator over an unbounded interval. As mentioned above, additional assumptions are necessary to control $\|\widetilde{f}_h\|_{L^1}$. We apply Lemma 4.2 assuming $\gamma \geqslant 4$, that is $\mathbb{E}[(\varepsilon_1^*)^4] < \infty$, and a polynomial decay of $|\varphi_\varepsilon'|$. The latter is a natural Mihlin-type condition in the context of Fourier multipliers. Note that φ_ε' exists if f_ε, the distribution of the measurement errors, has a first moment. In view of the analysis by Neumann and Reiß (2009) the moment assumption in particular implies uniform convergence of $\varphi_{\varepsilon,m}$.

To control the estimation error of $\widetilde{q}_{\tau,h}$, we follow the Z-estimator approach (cf. van der Vaart, 1998). Let $M(\eta)$ be the deterministic counterpart of $\widetilde{M}_h(\eta)$ defined in (4.4). It holds $M(q_\tau) = 0$ and we will see below that with high probability $\widetilde{M}_h(\widetilde{q}_{\tau,h}) = 0$. From the Taylor expansion $0 = \widetilde{M}_h(\widetilde{q}_{\tau,h}) = \widetilde{M}_h(q_\tau) + (\widetilde{q}_{\tau,h} - q_\tau)\widetilde{M}_h'(q_\tau^*)$ for some intermediate point q_τ^* between q_τ and $\widetilde{q}_{\tau,h}$, we obtain

$$\widetilde{q}_{\tau,h} - q_\tau = -\frac{\int_{-\infty}^{q_\tau}(\widetilde{f}_h(x) - f(x))\mathrm{d}x}{\widetilde{f}_h(q_\tau^*)}. \tag{4.9}$$

The following two propositions deal separately with the numerator and the denominator in this representation. The results are intrinsic to our analysis, but may also be of interest on their own. The first proposition deals with the numerator in (4.9) and establishes minimax rates of convergence for estimation of the distribution function with unknown error distributions. Note that the quotient in (4.9) might explode if $\widetilde{f}_h(q_\tau^*)$ becomes very small for large stochastic error. Excluding this event which has vanishing probability, we establish convergence rates as uniform $\mathcal{O}_P$-results, cf. definition (1.2).

Proposition 4.5. *Suppose that Assumption 4.A holds with $\ell = \langle\alpha\rangle + 1$ and let $h^*_{n,m} = (n \wedge m)^{-1/(2\alpha+2(\beta\vee 1/2)+1)}$. Then for any $\alpha \geqslant 1/2$, $\beta, R, r, \zeta > 0$ and $\gamma \geqslant 4$ we have uniformly over $f \in \mathcal{C}^\alpha(R, r, \zeta)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$ as $n \wedge m \to \infty$,*

$$\Big|\int_{-\infty}^{q_\tau} (\widetilde{f}_{h^*_{n,m}}(x) - f(x))\mathrm{d}x\Big| = \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}(\psi_{n\wedge m}(\alpha,\beta)),$$

where for $k \geqslant 1$

$$\psi_k(\alpha,\beta) := \begin{cases} k^{-1/2}, & \text{for } \beta \in (0, 1/2),\\ (\log k/k)^{1/2}, & \text{for } \beta = 1/2,\\ k^{-(\alpha+1)/(2\alpha+2\beta+1)}, & \text{for } \beta > 1/2. \end{cases} \tag{4.10}$$

Since the techniques to obtain Proposition 4.5 differ significantly from the literature on deconvolution with unknown error distribution, let us briefly sketch the proof: we apply a smooth truncation function a_s to decompose the error into

$$\begin{aligned}
&\int_{-\infty}^{q_\tau} (\widetilde{f}_h(x) - f(x))\mathrm{d}x\\
&\quad= \underbrace{\int_{-\infty}^{q_\tau} (K_h * f(x) - f(x))\mathrm{d}x}_{\text{deterministic error}} + \underbrace{\int_{-\infty}^{q_\tau} a_s(x+q_\tau)(\widetilde{f}_h(x) - K_h * f(x))\mathrm{d}x}_{\text{singular part of stochastic error}}\\
&\qquad+ \underbrace{\int_{-\infty}^{q_\tau} (1 - a_s(x+q_\tau))(\widetilde{f}_h(x) - K_h * f(x))\mathrm{d}x}_{\text{continuous part of stochastic error}}
\end{aligned} \tag{4.11}$$

with the usual notation $K_h(\cdot) = h^{-1}K(\cdot/h)$. The function a_s can be chosen such that it has compact support and satisfies $(\mathbb{1}_{(-\infty,0]} - a_s) \in C^\infty(\mathbb{R})$. Similar to the classical bias-variance trade-off, the deterministic error and singular part of the stochastic error will determine the rate. The continuous part, however, corresponds to the estimation error of a smooth (but not integrable) functional of the density. If the error distribution were known, it would be of order $n^{-1/2}$. For unknown errors we use Lemma 4.2, where our estimate of the operator norm of the random Fourier multiplier $\mathcal{F}^{-1}[\varphi_K(hu)/\varphi_{\varepsilon,m}(u)\mathbb{1}_{\{|\varphi_{\varepsilon,m}(u)|>m^{-1/2}\}}]$ is of order $\mathcal{O}_P(1 \vee (m^{-1/2}h^{-\beta-1}))$. This might be larger than the operator norm of the unknown deconvolution operator $\mathcal{F}^{-1}[1/\varphi_\varepsilon(u)]$ which is uniformly bounded. Yet, for $\alpha \geqslant 1/2$ the additional error that appears in the continuous part of stochastic error in (4.11) is negligible.

Next we like to understand the denominator of (4.9). Lounici and Nickl (2011) have proved uniform risk bounds for the deconvolution wavelet estimator on the whole real line for a known error distribution. On a bounded interval, which is sufficient for our purpose, uniform convergence of the deconvolution estimator $\widetilde{f}_h$ can be proved more elementarily. With $h_n = (\log n/n)^{1/(2\alpha+2\beta+1)}$ the following proposition yields the minimax rate $(\log n/n)^{\alpha/(2\alpha+2\beta+1)}$ in L^∞-loss (at least if $\frac{n}{\log n} \leqslant m$).

Proposition 4.6. *Grant Assumption 4.A with $\ell = \langle\alpha\rangle$. For any $\alpha, \beta, R, r, \zeta > 0$ and $\gamma \geqslant 0$ we have uniformly over $f \in \mathcal{C}^\alpha(R, r, \zeta)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$ as $n \wedge m \to \infty$,*

$$\sup_{x\in(-\zeta,\zeta)} |\widetilde{f}_h(x + q_\tau) - f(x + q_\tau)| = \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}\Big(h^\alpha + \big(\tfrac{\log n}{n} \vee \tfrac{1}{m}\big)^{1/2} h^{-\beta-1/2}\Big).$$

In particular, if $h = h_{n,m} \to 0$ and $(\frac{n}{\log n} \wedge m)h_{n,m}^{2\beta+1} \to \infty$ as $n \wedge m \to \infty$, $\widetilde{f}_{h_{n,m}}$ is a uniformly consistent estimator.

The two propositions above are the building blocks for the first main result of this paper announced in the following theorem. The constant preceding the rate depends only on the class parameters $\alpha, \beta, \gamma, R, r, \zeta$. The location parameter U_n can grow logarithmically to infinity as $n \to \infty$.

Theorem 4.7. *Let $\alpha \geqslant 1/2$, $\beta, R, r, \zeta > 0$ and $\gamma \geqslant 4$ and grant Assumption 4.A with $\ell = \langle\alpha\rangle + 1$. Let $\widetilde{q}_{\tau,h^*_{n,m}}$ be the quantile estimator defined in (4.4) associated with $h^*_{n,m} = (n \wedge m)^{-1/(2\alpha+2(\beta\vee1/2)+1)}$ and with $U_n \to \infty, U_n = \mathcal{O}(\log n)$. Then we have uniformly over $f \in \mathcal{C}^\alpha(R, r, \zeta, U_n)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$ as $n \wedge m \to \infty$,*

$$|\widetilde{q}_{\tau,h^*_{n,m}} - q_\tau| = \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}\big(\psi_{n\wedge m}(\alpha, \beta)\big)$$

where $\psi_\bullet(\alpha, \beta)$ is given in (4.10).

Using the methods of the proof of Theorem 4.7 and an additional application of Bernstein's concentration inequality, convergence rates for the uniform loss can be obtained, assuming regularity in a neighborhood of some interval of quantiles. For $0 < \tau_1 < \tau_2 < 1$ and $\alpha, R, r, \zeta, U_n > 0$ define

$$\begin{aligned}&\mathcal{C}^\alpha_\infty(\tau_1, \tau_2, R, r, \zeta, U_n)\\ &\quad := \Big\{f \in \mathcal{Q}(R) \Big| \text{ for all } \tau \in (\tau_1, \tau_2) : f \text{ has } \tau\text{-quantile } q_\tau \in [-U_n, U_n]\\ &\qquad\qquad \text{and } f \in C^\alpha([q_{\tau_1} - \zeta, q_{\tau_2} + \zeta], R), \inf_{\tau\in(\tau_1,\tau_2)} f(q_\tau) \geqslant r\Big\}.\end{aligned}$$

Theorem 4.8. *Let $\alpha \geqslant 1/2$, $\beta, R, r, \zeta > 0$ and $\gamma \geqslant 4$ and grant Assumption 4.A with $\ell = \langle\alpha\rangle + 1$. For $0 < \tau_1 < \tau_2 < 1$ and $\tau \in (\tau_1, \tau_2)$ let $\widetilde{q}_{\tau,h^*_{n,m}}$ be the quantile estimator defined in (4.4) associated with $h^*_{n,m} = (\frac{\log n}{n} \vee \frac{1}{m})^{1/(2\alpha+2(\beta\vee1/2)+1)}$ and with $U_n \to \infty, U_n = \mathcal{O}(\log n)$. Then we have uniformly over $f \in \mathcal{C}^\alpha_\infty(\tau_1, \tau_2, R, r, \zeta, U_n)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$ as $n \wedge m \to \infty$,*

$$\sup_{\tau\in(\tau_1,\tau_2)} |\widetilde{q}_{\tau,h^*_{n,m}} - q_\tau| = \mathcal{O}_{P,\mathcal{C}^\alpha_\infty\times\mathcal{D}^\beta}\big(\psi_{\frac{n}{\log n}\wedge m}(\alpha, \beta)\big)$$

where $\psi_\bullet(\alpha, \beta)$ is given in (4.10).

We finish this subsection by providing the minimax rates for estimating the distribution function and the quantiles for the case of known error distributions, restricting to pointwise loss. As above, the estimators are given by plugging in the classical density estimator

$$\widehat{f}_h(x) := \mathcal{F}^{-1}\Big[\frac{\varphi_n(u)\varphi_K(hu)}{\varphi_\varepsilon(u)}\Big](x), \quad x \in \mathbb{R}. \tag{4.12}$$

Corollary 4.9. *Let $\alpha, \beta, R, r, \zeta > 0$ and $\gamma \geqslant 0$ and suppose that the error distribution is known and $f_\varepsilon \in \mathcal{D}^\beta(R,\gamma)$. Let Assumption 4.A hold with $\ell = \langle\alpha\rangle + 1$. Let $\widehat{q}_{\tau,h}$ be the quantile estimator based on the density deconvolution estimator* (4.12) *associated with $h_n^* = n^{-1/(2\alpha+2(\beta\vee 1/2)+1)}$ and $U_n \to \infty, U_n = \mathcal{O}(\log n)$. Then we obtain uniformly over $f \in \mathcal{C}^\alpha(R, r, \zeta, U_n)$ as $n \to \infty$,*

$$\Big| \int_{-\infty}^{q_\tau} (\widehat{f}_{h_n^*}(x) - f(x))\mathrm{d}x \Big| = \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}(\psi_n(\alpha,\beta)),$$

$$|\widehat{q}_{\tau,h_n^*} - q_\tau| = \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}(\psi_n(\alpha,\beta)),$$

where $\psi_\bullet(\alpha,\beta)$ is given (4.10).

Here, we do not estimate the deconvolution operator and thus there is no additional error in terms of m. Consequently, we do not need a moment assumption on the error distribution and the convergence rates hold true for all $\alpha > 0$.

4.1.2. Lower bounds

In view of the lower bounds stated by Fan (1991b), in case $n \leq m$ the rates in Proposition 4.5 are optimal. Therefore, we focus on the case $m < n$. Using the error representation (4.9), the result carries over to quantile estimation. To provide a clear proof of the lower bound, we allow for a more general class of distributions of X_j, assuming only local assumptions. Using point measures, the estimation error of φ_ε does not profit from the decay of the characteristic function of X_j. One could also consider the case of bounded densities f and choose alternatives in the proof whose Fourier transforms decay arbitrarily slowly, but this would require far more technical arguments.

We define for $\alpha, R, r, \zeta > 0$ and some interval $I \subseteq \mathbb{R}$

$$\begin{aligned}
\widetilde{\mathcal{C}}^\alpha(R,r,I) &:= \Big\{ F \text{ c.d.f.} \Big| F \text{ has on } I \text{ a Lebesgue density } f \in C^\alpha(I,R) \\
&\qquad\qquad \text{and } \inf_{x\in I} f(x) \geqslant r \Big\}, \\
\widetilde{\mathcal{C}}^\alpha(R,r,\zeta) &:= \Big\{ F \text{ c.d.f.} \Big| F \text{ has a } \tau\text{-quantile } q_\tau \in \mathbb{R} \text{ and } F \in \widetilde{\mathcal{C}}^\alpha(R,r,[q_\tau-\zeta, q_\tau+\zeta]) \Big\}.
\end{aligned}$$

Theorem 4.10. *Suppose that $Y_1,\dots,Y_n$ and $\varepsilon_1^*,\dots,\varepsilon_m^*$ are independent. Let $q \in \mathbb{R}$ and $\alpha,\beta,R,r,\zeta > 0, \gamma \geqslant 0$. Then for any $C > 0$ there is some $\delta > 0$ such that*

$$\inf_{\bar{F}_{n,m}} \sup_{\widetilde{\mathcal{C}}^\alpha(R,r,[q-\zeta,q+\zeta])} \sup_{\mathcal{D}^\beta(R,\gamma)} P\Big(|\bar{F}_{n,m}(q) - F(q)| > C(n\wedge m)^{-(\alpha+1)/(2\alpha+(2\beta)\vee 1+1)}\Big) \geqslant \delta,$$

$$\inf_{\bar{q}_{\tau,n,m}} \sup_{\widetilde{\mathcal{C}}^\alpha(R,r,\zeta)} \sup_{\mathcal{D}^\beta(R,\gamma)} P\Big(|\bar{q}_{\tau,n,m} - q_\tau| > C(n\wedge m)^{-(\alpha+1)/(2\alpha+(2\beta)\vee 1+1)}\Big) \geqslant \delta,$$

where the infima are taken over all estimators $\bar{F}_{n,m}$ and $\bar{q}_{\tau,n,m}$, respectively.

This lower bound implies that the rates in Proposition 4.5 and Theorem 4.7 are minimax optimal, except for the case $\beta = 1/2$ where they deviate by a logarithmic factor.

4.1.3. Discussion and extension

The results of this section show that estimating the distribution function by integrating a density deconvolution estimator is a minimax optimal procedure and under the local Hölder condition the rates are determined by $n \wedge m$. In that point our results differ completely from previous studies. Assuming α-Sobolev regularity of f, the RMSE of the kernel density estimator by Neumann (1997) is of order $\mathcal{O}(n^{-\alpha/(2\alpha+\beta+1)} + m^{-((\alpha/\beta)\wedge 1)})$. Since the error in estimating φ_ε is reduced by the decay of the characteristic function φ of X_j, the risk is of much smaller order in m. Assuming local regularity on f only, $\mathcal{F}f$ can decay arbitrarily slowly such that this reduction effect may not occur. Note that assuming global Sobolev regularity would improve also the convergence rate of the plug-in estimator.

Interestingly, the dependence on n and m is not completely symmetric. As an intrinsic property of the uniform loss, the convergence rates are typically by a logarithmic factor slower than for pointwise loss. Yet, in Proposition 4.6 and Theorem 4.8 this payment for uniform convergence affects only the estimation of φ and thus the rate is determined by $\frac{\log n}{n} \vee \frac{1}{m}$.

Although the focus of this paper is on ordinary smooth error distributions, a generalization to supersmooth errors is worth mentioning. Let us sketch this case of exponentially decaying φ_ε. Supposing $\mathbb{E}[|\varepsilon_k^*|^4] < \infty$ and $|\varphi_\varepsilon(u)|^{-1} \lesssim e^{\gamma_0|u|^\beta}$ as well as $|\varphi_\varepsilon'(u)| \lesssim e^{-\gamma_1|u|^\beta}, u \in \mathbb{R}$, for some $\beta > 0$ and $\gamma_0 \geqslant \gamma_1 > 0$, we obtain analogously to Lemma 4.2 for sufficiently small $c, \gamma > 0$ and for the bandwidth $h_m^* = c(\log m)^{-1/\beta}$

$$\Big\| \mathcal{F}^{-1}\Big[\frac{\varphi_K(h_m^* u)}{\varphi_{\varepsilon,m}(u)}\Big] * \psi \Big\|_{C^s} \mathbb{1}_{B_\varepsilon(h_m^*)} \leqslant \mathcal{E}_{h_m^*} \|\psi\|_{C^s} \quad \text{where} \quad \mathcal{E}_h = \mathcal{O}_P\big(1 \vee e^{\gamma h^{-\beta}}\big)$$

for any $s \geqslant 0$ and for any $\psi \in C^s(\mathbb{R})$. In other words, $\varphi_K(hu)/\varphi_{\varepsilon,m}(u)$ is a random Fourier multiplier on Hölder spaces with exponentially increasing operator norm on the event $B_\varepsilon(h)$. Following the lines of the proof of Proposition 4.5, one sees that the singular as well as the continuous part of the stochastic error in (4.11) are of the order $\mathcal{O}_P((n \wedge m)^{-1/2} e^{\gamma h^{-\beta}})$. Combined with the estimate for the deterministic error, the choice $h_{n,m}^* = c(\log(n \wedge m))^{-1/\beta}$ yields for $f \in \mathcal{C}^\alpha(R, r, \zeta)$

$$\Big| \int_{-\infty}^{q_\tau} (\widetilde{f}_{h_n^*}(x) - f(x))\mathrm{d}x \Big| = \mathcal{O}_P\big((\log(n \wedge m))^{-(\alpha+1)/\beta}\big).$$

Note that for $n \leqslant m$ this is the minimax rate for distribution function estimation as given in Fan (1991b). Therefore, also for supersmooth error distributions the integral domain does not need to be truncated to estimate the distribution function via the plug-in approach.

4.2. Adaptive estimation

The choice of the bandwidth h is crucial in applications. Therefore, we develop a fully data-driven procedure to determine a good bandwidth. We follow the approach initiated by Lepski (1990). More precisely, we use the version iterated confidence intervals, cf. Goldenshluger and Nemirovski (1997). For simplicity, we suppose $n = m$ and focus on the pointwise loss in this section.

Let us consider the family of estimators $\{\widetilde{q}_{\tau,h}, h \in \mathcal{B}_n\}$ where $\widetilde{q}_{\tau,h}$ is defined in (4.4) and $\mathcal{B}_n$ is a finite set of bandwidths. In view of the error representation (4.9), it is important that $\widetilde{f}_h(\widetilde{q}_{\tau,h})$ is a consistent estimator of $f(q_\tau)$ for all $h \in \mathcal{B}_n$. Therefore, conditions on the bandwidth as in Proposition 4.6 are necessary for the entire set $\mathcal{B}_n$. These depend on the true but unknown degree of ill-posedness β and on α. We keep to the assumption $\alpha > 1/2$ such that the additional error due to bounding the random Fourier multiplier is negligible. Note that the lower bound for the bandwidth is not determined by the variance of the quantile estimator itself but by the variance of the density estimator and the minimal smoothing which results from $\alpha > 1/2$.

Inspired by Comte and Lacour (2011), we propose the following construction of a feasible set $\mathcal{B}_n$: For some $L > 1$ define

$$h_{n,j} := n^{-1}L^j \text{ for } j = 0, \ldots, N_n \quad \text{where} \quad N_n \in \mathbb{N} \text{ satisfies } n^{-1}L^{N_n} \sim (\log n)^{-3}.$$

Choosing

$$\widetilde{\jmath}_n := \min\Big\{j = 0, \ldots, N_n : \frac{1}{2} \leqslant \Big(\frac{\log n}{n}\Big)^{1/2} \int_{-1/h_{n,j}}^{1/h_{n,j}} \frac{\mathbb{1}_{\{|\varphi_{\varepsilon,m}(u)| \geqslant m^{-1/2}\}}}{|\varphi_{\varepsilon,m}(u)|} \mathrm{d}u \leqslant 1\Big\}, \quad (4.13)$$

the bandwidth set is given by

$$\mathcal{B}_n := \{h_{n,\widetilde{\jmath}_n}, \ldots, h_{n,N_n}\}. \quad (4.14)$$

Lemma 4.11. *Let (Y_j) and (ε_k^*) be distributed according to $f \in \mathcal{C}^\alpha(R, r, \zeta)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$ with $\alpha > 1/2, \beta > 0$. With probability converging to one $\mathcal{B}_n$ from (4.14) satisfies:*

(i) $\mathcal{B}_n$ consists of a monotone increasing sequence of bandwidths such that $h_{n,j+1}/h_{n,j}$ is uniformly bounded in $j = \widetilde{\jmath}_n, \ldots, N_n$ and $n \geqslant 1$.

(ii) For $n \to \infty$ we have $N_n \lesssim \log n, (\log n)^2 h_{n,N_n} \to 0$ and $nh_{n,\widetilde{\jmath}_n}^{2\beta+2} \to \infty$.

(iii) The optimal bandwidth $h_n^* = n^{-1/(2\alpha+2(\beta\vee 1/2)+1)}$ is contained in $[h_{n,\widetilde{\jmath}_n}, h_{n,N_n}]$.

Given the bandwidth set, the adaptive estimator is obtained by selection from the family of estimators $\{\widetilde{q}_{\tau,h}, h \in \mathcal{B}_n\}$. As proposed by Lepski (1990) the adaptive choice should mimic the trade-off between deterministic error and stochastic error. We select the largest bandwidth such that the intersection of all confidence sets, which corresponds to smaller bandwidths, is non-empty. As discussed above it is sufficient to consider the singular part of the stochastic error in (4.11) only. To estimate the variance of $\widetilde{q}_{\tau,h}$ corresponding to the latter, we define for some $\delta > 0$

$$\widetilde{\Sigma}_h := \frac{(2\sqrt{2} + \delta)\sqrt{\log\log n}\max_{\mu \geqslant h} \widetilde{\sigma}_{\mu,X} + (\delta \log n)^3 \max_{\mu \geqslant h} \widetilde{\sigma}_{\mu,\varepsilon}}{|\widetilde{f}_h(\widetilde{q}_{\tau,h})|}, \quad (4.15)$$

with the truncation function a_s from decomposition (4.11) and

$$\widetilde{\sigma}_{h,X}^2 = \frac{1}{n^2}\sum_{j=1}^n \Big(\int_{-\infty}^0 a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_{\varepsilon,m}(u)}\Big](x + \widetilde{q}_{\tau,h})\mathrm{d}x\Big)^2 \quad \text{and} \quad (4.16)$$

$$\widetilde{\sigma}_{h,\varepsilon}^2 = \frac{1}{4\pi^2 m}\int_{-1/h}^{1/h} |\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)}\Big|^2 \mathrm{d}u \int_{-1/h}^{1/h} |\varphi_K(hu)|\Big|\frac{\mathcal{F}\,a_s(u)}{\varphi_{\varepsilon,m}(u)}\Big|^2 \mathrm{d}u. \quad (4.17)$$

The parameter δ has minor influence and should be chosen close to zero. Note that we apply a monotonization in the numerator of $\widetilde{\Sigma}_h$ by taking maxima of $\widetilde{\sigma}_{\mu,X}$ and $\widetilde{\sigma}_{\mu,\varepsilon}$, respectively. Define for any $h \in \mathcal{B}_n$

$$\mathcal{U}_h := [\widetilde{q}_{\tau,h} - \widetilde{\Sigma}_h, \widetilde{q}_{\tau,h} + \widetilde{\Sigma}_h]. \tag{4.18}$$

The adaptive estimator is given by

$$\widetilde{q}_\tau := \widetilde{q}_{\tau,\widetilde{h}_n^*} \quad \text{with} \quad \widetilde{h}_n^* := \max\Big\{h \in \mathcal{B}_n \,\Big|\, \bigcap_{\mu \le h, \mu \in \mathcal{B}_n} \mathcal{U}_\mu \neq \emptyset\Big\}. \tag{4.19}$$

Note that $\widetilde{h}_n^*$ is well defined since the intersection in (4.19) is non-empty for $h = h_{n,\widetilde{\jmath}_n}$. The following theorem shows that this estimator achieves the minimax rate up to a logarithmic factor. The proof relies on a comparison with an oracle-type choice of the bandwidth. All ingredients, though, have to be estimated and the dependence between Y_j and ε_k^* requires special attention.

Theorem 4.12. *Let $n = m$ and $\alpha > 1/2$, $\beta, R, r, \zeta > 0, \gamma \geqslant 4$ and grant Assumption 4.A with $\ell \geqslant \langle\alpha\rangle + 1$. Then the estimator $\widetilde{q}_\tau$ as defined in (4.19) with $\mathcal{B}_n$ from* (4.14) *satisfies uniformly over $f \in \mathcal{C}^\alpha(R, r, \zeta, U_n)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$ as $n \to \infty$,*

$$|\widetilde{q}_\tau - q_\tau| = \mathcal{O}_{P,\mathcal{C}^\alpha \times \mathcal{D}^\beta}\Big(\psi_{n(\delta \log n)^{-6}}(\alpha, \beta)\Big),$$

where $\psi_\bullet(\alpha, \beta)$ is given in (4.10).

As the theorem shows, the adaptive method achieves the minimax rate up to a logarithmic factor. This additional loss is dominated by the stochastic error which is due to the estimation of φ_ε. Since Y_j and ε_k^* are not independent, we have to bound the stochastic error of $\widetilde{q}_{\tau,h}$ in a way that separates the error terms coming from the estimation of φ and φ_ε, respectively. Estimating the remaining parts, we lose the factor $(\delta \log n)^6$, which appears not to be optimal. To improve the rate slightly, $\delta = \delta_n$ could be chosen as a null sequence provided $\delta_n(\log n)^{1/2} \to \infty$. In the case where the error density is known, we can achieve the better rate $\psi_{n/\log\log n}(\alpha, \beta)$. The $\log\log n$-factor is the additional payment for $\mathcal{O}_P$-adaptivity, which is known to be unavoidable for a bounded loss function in standard regression, cf. Spokoiny (1996). For estimating the distribution function, an analogous result can be obtained, but is omitted.

4.3. Numerical results

4.3.1. Simulation study

We illustrate the implementation of the adaptive estimation procedure of Section 4.2. Our small simulation study serves as a proof of viability of the proposed method.

We run 1000 Monte Carlo simulations for four experimental setups. The sample size is set to $n = 1000$ and the external sample of the directly observed error is set to $m = 1000$ as well (here the external sample is independent of the main one). We consider the gamma distributions $\Gamma(1, 1)$ and $\Gamma(2, 1)$. Note that the shape k of the gamma distribution determines the Sobolev smoothness of the density while the density

RMSE	$k = 1, \beta = 2$	$k = 2, \beta = 2$	$k = 1, \beta = 4$	$k = 2, \beta = 4$
$\tau = 0.1$	0.377 (1.028)	0.188 (0.767)	0.392 (1.758)	0.253 (1.427)
$\tau = 0.2$	0.175 (0.452)	0.094 (0.353)	0.203 (0.920)	0.158 (0.731)
$\tau = 0.3$	0.072 (0.178)	0.084 (0.155)	0.124 (0.433)	0.148 (0.344)
$\tau = 0.4$	0.114 (0.051)	0.103 (0.066)	0.154 (0.102)	0.155 (0.095)
$\tau = 0.5$	0.170 (0.174)	0.115 (0.150)	0.200 (0.234)	0.160 (0.206)
$\tau = 0.6$	0.200 (0.317)	0.108 (0.257)	0.235 (0.501)	0.163 (0.424)
$\tau = 0.7$	0.185 (0.461)	0.099 (0.372)	0.238 (0.778)	0.170 (0.644)
$\tau = 0.8$	0.119 (0.630)	0.144 (0.505)	0.200 (1.084)	0.236 (0.893)
$\tau = 0.9$	0.229 (0.850)	0.285 (0.676)	0.379 (1.486)	0.538 (1.219)

Table 4.1.: Empirical root mean square error (RMSE) of the adaptive and naive (in parenthesis) estimators for estimating q_τ based on 1000 Monte Carlo simulations with $n = m = 1000$.

is smooth away from the origin. For the error distribution we consider the standard Laplace distribution ($\beta = 2$) and the convolution of a standard Laplace with itself ($\beta = 4$).

The target quantiles of interest are q_τ with $\tau = 0.1, 0.2, \dots, 0.9$. In the real data example in the next subsection we compare the adaptive estimator to the "naive" quantile estimator given by the empirical quantiles of the observations Y. Therefore we have also applied the naive estimator in the simulations. The results of this simulation study are given in Table 4.1. We can see that the results support the theory: The empirical root mean squared error (RMSE) is higher for $\beta = 4$ than for $\beta = 2$. Also, we can see that in most cases the RMSE is lower for $k = 2$ than for $k = 1$ since the gamma distribution with larger shape parameter is smoother in our context. At the tails our estimation method is significantly better than the naive estimator. Near the median the naive estimator profits from the symmetric error distribution and thus achieves comparable (or even slightly better) results.

4.3.2. Real data example

High blood pressure is a direct cause of serious cardiovascular disease (Kannel et al. (1995)) and determining reference values for physicians is important. In particular, estimating percentiles of systolic and diastolic blood pressure by sex, race or ethnicity, age, etc. is of substantial interest. Blood pressure is known to be measured with additional error which needs to be addressed in its analysis (see e.g., Frese et al. (2011)). Therefore, measurement errors should be taken into account, otherwise quantile estimates based on the observed blood pressure measurements would be biased.

We illustrate our method using data from the Framingham Heart Study (Carroll et al. (2006)). This study consists of a series of exams taken two years apart where systolic blood pressure (SBP) measurements of $1,615$ men aged $31 - 65$ were taken. These data were used as an illustration for density estimation in deconvolution by Stirnemann et al. (2012) and for distribution function estimation by Dattner and Reiser (2013). We

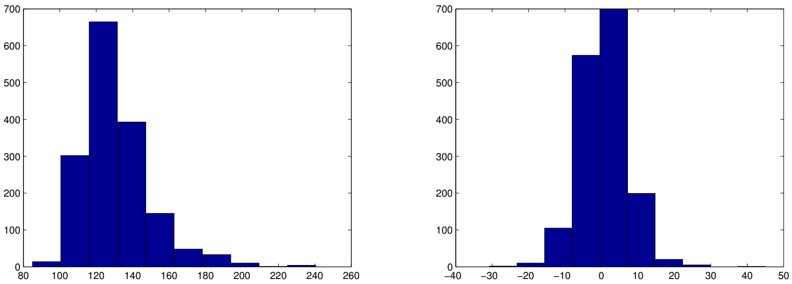

Figure 4.1.: Average systolic blood pressure Y' *(left)* and the errors ε^* *(right)* over the two measurements from the two visits of $1,615$ men aged $31-65$ from the Framingham Heart Study.

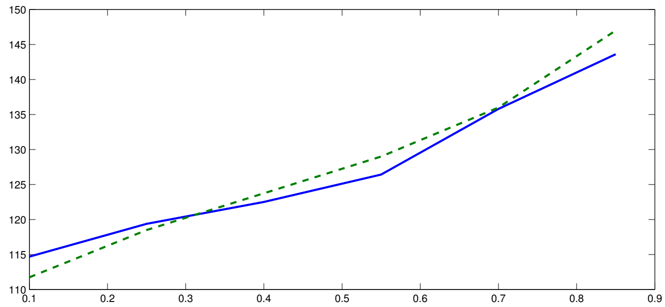

Figure 4.2.: Quantiles estimates for systolic blood pressure of $1,615$ men aged $31-65$ from the Framingham Heart Study. Solid line for the adaptive estimator and dashed line for the naive estimator.

denote by $Y_{j,1}$ and $Y_{j,2}$ the two repeated measures of SBP for each individual j at two different exams and denote by X_j the long-term average SBP of individual j. Then we model that

$$Y_{j,1} = X_j + \varepsilon_{j,1}, \qquad Y_{j,2} = X_j + \varepsilon_{j,2},$$

for individuals $j = 1, ..., n$. Following Carroll et al. (2006), we use the average of the two exams $Y_j' = (Y_{j,1} + Y_{j,2})/2$, so that the model in our case is

$$Y_j' = X_j + \varepsilon_j', \tag{4.20}$$

where $\varepsilon_j' = (\varepsilon_{j,1} + \varepsilon_{j,2})/2$.

Taking advantage of the repeated measurements, we can avoid parametric assumptions regarding the distribution of the errors. The only assumption we will make is that the distribution of the measurement error is symmetric around zero and does not vanish. We then set $\varepsilon_j^* = (Y_{j,1} - Y_{j,2})/2$ and note that under the symmetry assumption it is distributed as ε_j'. We emphasize the fact that our theoretical results do not require that the sample ε_j^* must be independent from that of the Y_j'.

Histograms of Y' and ε^* are presented in Figure 4.1. The resulting adaptive and naive quantiles estimates are displayed in Figure 4.2. We can see certain differences between the naive and adaptive estimates which might result in important implications for medical research, but here we do not aim at pursuing a more detailed statistical analysis.

4.4. Proofs

4.4.1. Proofs for Section 4.1

For a better readability we assume throughout $\beta \neq 1/2$. In the special case $\beta = 1/2$ the order of the stochastic error will be $(\log n/n)^{1/2}$ which can be easily seen below in the bounds (4.32) and (4.34). The subscript n at the bandwidth will be omitted.

Since $1/\varphi_{\varepsilon,m}$ might explode for large stochastic errors we need the following lemma.

Lemma 4.13. *Suppose $\mathbb{E}[|\varepsilon_k^*|^\delta] < \infty$ for some $\delta > 0$. Let $T_m \to \infty$ be an increasing sequence satisfying $m^{1/2} \inf_{u\in[-T_m,T_m]} |\varphi_\varepsilon(u)| \gtrsim (\log T_m)^2$, then for any $p < 2$*

$$P\Big(\inf_{u\in[-T_m,T_m]} |\varphi_{\varepsilon,m}(u)| < m^{-1/2}(\log T_m)^p\Big) = o(1) \quad as \quad m \to \infty.$$

Proof. The triangle inequality, the assumption on T_m and Markov's inequality yield for m as well as T_m large enough

$$\begin{aligned}
&P\Big(\inf_{u\in[-T_m,T_m]} |\varphi_{\varepsilon,m}(u)| < m^{-1/2}(\log T_m)^p\Big)\\
\leqslant &P\Big(\sup_{u\in[-T_m,T_m]} |\varphi_\varepsilon(u) - \varphi_{\varepsilon,m}(u)| > \inf_{u\in[-T_m,T_m]} |\varphi_\varepsilon(u)| - m^{-1/2}(\log T_m)^p\Big)\\
\lesssim &\frac{2}{(\log T_m)^2}\,\mathbb{E}\Big[\sup_{u\in[-T_m,T_m]} m^{1/2}|\varphi_\varepsilon(u) - \varphi_{\varepsilon,m}(u)|\Big].
\end{aligned}$$

Noting $\mathbb{1}_{[-T_m,T_m]}(u) \leqslant w(u)/w(T_m)$ for $w(u) := (\log(e + |u|))^{-1/2-\eta}$ for some $\eta \in (0, 1/2)$, the above display can be bounded by

$$\frac{2}{w(T_m)(\log T_m)^2}\,\mathbb{E}\Big[\sup_{u\in\mathbb{R}} m^{1/2} w(u)|\varphi_\varepsilon(u) - \varphi_{\varepsilon,m}(u)|\Big] \lesssim (\log T_m)^{-3/2+\eta} \tag{4.21}$$

where the expectation is bounded by applying Theorem 4.1 in Neumann and Reiß (2009). □

To ensure consistency of the density estimator, the bandwidth satisfies usually $(n \wedge m)h^{2\beta+1} \to \infty$ and is of polynomial order in n, m. This implies $\inf_{u\in[-1/h,1/h]} |\varphi_\varepsilon(u)| \gtrsim m^{-1/2}|\log h|^2$ for $f \in \mathcal{D}^\beta(R,\gamma), \gamma > 0$, and thus Lemma 4.13 can be applied to $T_m = 1/h$. Hence, under this conditions on h the probability of the event $B_\varepsilon(h)$, defined in (4.6), tends to one. In that case, it suffices to control terms on $B_\varepsilon(h)$. Frequently, the weaker estimate $|\varphi_{\varepsilon,m}(u)| \geqslant m^{-1/2}$ for $|u| \leqslant 1/h$ will be enough. In particular,

$$\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\mathbb{1}_{\{|\varphi_{\varepsilon,m}(u)|\geq m^{-1/2}\}} = \frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)} \quad \text{on } B_\varepsilon(h).$$

Proof of Lemma 4.1

On $B_\varepsilon(h)$ we have by continuity of the characteristic functions and the properties of the kernel that $(\frac{\varphi_n \varphi_K(h\bullet)}{\varphi_{\varepsilon,m}})^{(l)} \in L^2(\mathbb{R})$, for $l = 0, 1$, which implies $(\varphi_n \varphi_K(h\bullet))/\varphi_{\varepsilon,m} \in B^1_{2,2}(\mathbb{R}) \subseteq B^{1/2}_{2,1}(\mathbb{R})$ (for Besov spaces see Definition 2.3.1(2) in Triebel (2010)). Hence, Corollary 3.2 by Girardi and Weis (2003) yields $\widetilde{f}_h \in L^1(\mathbb{R})$ verifying that (4.5) is well defined on the event $B_\varepsilon(h)$.

On $B_\varepsilon(h)$ we have moreover $\lim_{\eta\to-\infty} \int_{-\infty}^{\eta} \widetilde{f}_h(x)\mathrm{d}x = 0$, by integrability of $\widetilde{f}_h$, and $\int_{-\infty}^{\infty} \widetilde{f}_h(x)\mathrm{d}x = \mathcal{F}[\widetilde{f}_h](0) = \varphi_n(0)\varphi_K(0)/\varphi_{\varepsilon,m}(0) = 1$. Applying the bound $\|\widetilde{f}_h\|_\infty \leqslant \|\varphi_K(hu)/\varphi_{\varepsilon,m}(u)\|_{L^1} < \infty$, we conclude that $\eta \mapsto \int_{-\infty}^{\eta} \widetilde{f}_h(x)\mathrm{d}x$ continuous and $[0,1]$ is contained in its range. □

Proof of Lemma 4.2

Note that the assumption on φ_ε imply $|(\varphi_\varepsilon^{-1})'(u)| \lesssim (1+|u|)^{\beta-1}$ as well as $|\varphi_\varepsilon^{-1}(u)| \lesssim (1+|u|)^\beta, u \in \mathbb{R}$. We define the random Fourier multiplier

$$\psi(u) := (1+iu)^{-\beta} \frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}, \quad u \in \mathbb{R}\,.$$

On $B_\varepsilon(h)$, as defined in (4.6), we will check Hörmander type conditions and derive an upper bound for the operator norm of $\psi(u)$. Hence, we have to determine a suitable constant $A_\psi > 0$ satisfying

$$\begin{aligned} \max_{l\in\{0,1\}} \Big(\int_{[-2,2]} |\psi^{(l)}(u)|^2 \mathrm{d}u \Big)^{1/2} &\leqslant A_\psi \quad \text{and} \\ \max_{l\in\{0,1\}} \sup_{T\in[1,\infty)} T^{l-1/2} \Big(\int_{T\leqslant|u|\leqslant 4T} |\psi^{(l)}(u)|^2 \mathrm{d}u \Big)^{1/2} &\leqslant A_\psi. \end{aligned} \tag{4.22}$$

To find A_ψ, we note that

$$\frac{1}{|\varphi_{\varepsilon,m}(u)|^p} \leqslant \frac{p}{|\varphi_\varepsilon(u)|^p} + \frac{p|\varphi_{\varepsilon,m}(u) - \varphi_\varepsilon(u)|^p}{|\varphi_\varepsilon(u)\varphi_{\varepsilon,m}(u)|^p}, \quad \text{for } p \in \{1,2\} \tag{4.23}$$

and thus on $B_\varepsilon(h)$

$$\frac{1}{|\varphi_{\varepsilon,m}(u)|} \leqslant \frac{1+\Delta_m(u)}{|\varphi_\varepsilon(u)|}, \qquad \Delta_m(u) := \frac{m^{1/2}}{|\log h|^{3/2}} |\varphi_{\varepsilon,m}(u) - \varphi_\varepsilon(u)|,$$

By the assumptions on φ_ε and K we conclude

$$|\psi(u)| \leqslant \frac{|\varphi_K(hu)|(1+\Delta_m(u))}{(1+u^2)^{\beta/2}|\varphi_\varepsilon(u)|} \lesssim (1+\Delta_m(u))\mathbb{1}_{[-1/h,1/h]}(u). \tag{4.24}$$

Concerning the derivative, we estimate $h \leqslant 2(1+|u|)^{-1}$ for $|u| \leqslant 1/h$ and $h < 1/2$ and consequently by $|\varphi'_\varepsilon(u)/\varphi_\varepsilon(u)| \lesssim (1+|u|)^{-1}$

$$
\begin{aligned}
|\psi'(u)| \leqslant{}& (\beta+1)(1+u^2)^{-(\beta+1)/2}\Big|\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big| + h(1+u^2)^{-\beta/2}\Big|\frac{\varphi'_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big| \\
&+ (1+u^2)^{-\beta/2}\Big|\frac{\varphi'_{\varepsilon,m}(u)}{\varphi_{\varepsilon,m}(u)}\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big| \\
\lesssim{}& \frac{|\psi(u)|}{1+|u|} + |\psi(u)|\Big|\frac{\varphi'_{\varepsilon,m}(u)}{\varphi_{\varepsilon,m}(u)}\Big| \\
\lesssim{}& (1+\Delta_m(u))\Big(\frac{1}{1+|u|} + (1+\Delta_m(u))\Big|\frac{\varphi'_{\varepsilon,m}(u)}{\varphi_\varepsilon(u)}\Big|\Big)\mathbb{1}_{[-1/h,1/h]}(u) \\
\lesssim{}& (1+\Delta_m(u))\Big(\frac{1}{1+|u|} + (1+\Delta_m(u))(1+|u|)^\beta|\varphi'_{\varepsilon,m}(u) - \varphi'_\varepsilon(u)|\Big)\mathbb{1}_{[-1/h,1/h]}(u) \\
\lesssim{}& \frac{(1+\Delta_m(u))^2}{1+|u|}\Big(1 + (1+|u|)^{\beta+1}|\varphi'_{\varepsilon,m}(u) - \varphi'_\varepsilon(u)|\Big)\mathbb{1}_{[-1/h,1/h]}(u).
\end{aligned} \tag{4.25}
$$

With these bounds at hand we can show (4.22). For $l = 0$ the estimate (4.24) and $1/T \lesssim (1+|u|)^{-1}$ for $|u| \leqslant 4T$ yield

$$
\begin{aligned}
\int_{-2}^{2}|\psi(u)|^2\mathrm{d}u &\lesssim \int_{-2}^{2}(1+\Delta_m^2(u))\mathbb{1}_{[-1/h,1/h]}(u)\mathrm{d}u, \\
\frac{1}{T}\int_{T\leqslant|u|\leqslant 4T}|\psi(u)|^2\mathrm{d}u &\lesssim \frac{1}{T}\int_{T\leqslant|u|\leqslant 4T}(1+\Delta_m^2(u))\mathbb{1}_{[-1/h,1/h]}(u)\mathrm{d}u \\
&\lesssim 1 + \int_{-1/h}^{1/h}(1+|u|)^{-1}\Delta_m^2(u)\mathrm{d}u,
\end{aligned}
$$

for h small enough. Hence, the conditions (4.22) for $l = 0$ are satisfied for A_ψ of the order $(1+\int_{-1/h}^{1/h}(1+|u|)^{-1}\Delta_m^2(u)\mathrm{d}u)^{1/2}$. For $l = 1$ we verify by (4.25) and $T \leqslant (1+|u|)$ for $|u| > T$

$$
\begin{aligned}
&\int_{-2}^{2}|\psi'(u)|^2\mathrm{d}u \lesssim \int_{-2}^{2}(1+\Delta_m^4(u))\big(1+(1+|u|)^{2\beta+2}|\varphi'_{\varepsilon,m}(u)-\varphi'_\varepsilon(u)|^2\big)\mathrm{d}u \quad \text{and} \\
&T\int_{T\leqslant|u|\leqslant 4T}|\psi'(u)|^2\,\mathrm{d}u \\
&\quad\lesssim \int_{T\leqslant|u|\leqslant 4T}T(1+|u|)^{-2}\mathrm{d}u \\
&\qquad + \int_{-1/h}^{1/h}\Big(\frac{\Delta_m^4(u)}{1+|u|} + (1+\Delta_m^4(u))(1+|u|)^{2\beta+1}|\varphi'_{\varepsilon,m}(u)-\varphi'_\varepsilon(u)|^2\Big)\mathrm{d}u \\
&\quad\lesssim 1 + \int_{-1/h}^{1/h}\Big(\frac{\Delta_m^4(u)}{1+|u|} + (1+\Delta_m^4(u))(1+|u|)^{2\beta+1}|\varphi'_{\varepsilon,m}(u)-\varphi'_\varepsilon(u)|^2\Big)\mathrm{d}u.
\end{aligned}
$$

Therefore, we find a constant $A' > 0$, depending only on R, β, such that (4.22) holds for

$$
A_\psi := A'\Big(1 + \int_{-1/h}^{1/h}\Big(\frac{\Delta_m^2(u)+\Delta_m^4(u)}{1+|u|}
$$

$$+ (1+\Delta_m^4(u))(1+|u|)^{2\beta+1}|\varphi'_{\varepsilon,m}(u)-\varphi'_\varepsilon(u)|^2\Big)\mathrm{d}u\Big)^{1/2}. \quad (4.26)$$

The conditions (4.22) imply that ψ is indeed a Fourier multiplier on $B_\varepsilon(h)$ and thus by Theorem 4.8 and Corollary 4.13 by Girardi and Weis (2003) with $p = 2$, $l = 1$ there is a universal constant $C > 0$ such that for all $\eta > 0$ and $f \in C^{s+\beta+\eta}(\mathbb{R})$

$$\Big\| \mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big] * f\Big\|_{C^s} = \Big\| \mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\mathcal{F} f\Big]\Big\|_{C^s} \leqslant CA_\psi \Big\| \mathcal{F}^{-1}\left[(1+iu)^\beta \mathcal{F} f\right]\Big\|_{C^{s+\eta}}.$$

Choosing $\eta > 0$ such that $s+\beta+\eta, s+\eta \notin \mathbb{N}$, the Fourier multiplier $(1+iu)^\beta$ induces an isomorphism from $C^{s+\beta+\eta}(\mathbb{R})$ onto $C^{s+\eta}(\mathbb{R})$ (Triebel, 2010, Thm. 2.3.8). Hence, there is another universal constant $C' > 0$ such that the first assertion of the lemma follows:

$$\Big\| \mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\mathcal{F} f\Big]\Big\|_{C^s} \leqslant \mathcal{E}_h \|f\|_{C^{s+\beta+\eta}} \quad \text{with} \quad \mathcal{E}_h := C'A_\psi.$$

To bound $\mathcal{E}_h$, we apply Markov's inequality on A_ψ from (4.26). The inequality by Rosenthal (1970) yields

$$\sup_{u\in\mathbb{R}} \mathbb{E}\Big[m^{p/2}\Big|\varphi^{(l)}_{\varepsilon,m}(u) - \varphi^{(l)}_\varepsilon(u)\Big|^p\Big] < \infty$$

for $l = 0$ and $p \in \mathbb{N}$ as well as $l = 1$ and $p \in \{1,\dots,4\}$. Combined with the Markov inequality and Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
&P\Big(B_\varepsilon(h) \cap \Big\{\mathcal{E}_h > \frac{c^{1/2}}{m^{1/2}h^{\beta+1}\wedge 1}\Big\}\Big) \leqslant c^{-1}(mh^{2\beta+2}\wedge 1)\,\mathbb{E}\left[\mathcal{E}_h^2 \mathbb{1}_{B_\varepsilon(h)}\right] \\
&\quad \lesssim \frac{1}{c}(mh^{2\beta+2}\wedge 1)\Big(1 + \int_{-1/h}^{1/h}\Big((1+|u|)^{-1}\,\mathbb{E}[\Delta_m^2(u) + \Delta_m^4(u)] \\
&\qquad + \mathbb{E}\Big[(1+\Delta_m^4(u))(1+|u|)^{2\beta+1}|\varphi'_{\varepsilon,m}(u) - \varphi'_\varepsilon(u)|^2\Big]\Big)\mathrm{d}u\Big) \\
&\quad \lesssim \frac{mh^{2\beta+2}\wedge 1}{c}\Big(1 + \frac{1}{|\log h|^3}\int_{-1/h}^{1/h}\frac{\mathrm{d}u}{1+|u|} + \frac{1}{m}\int_{-1/h}^{1/h}(1+|u|)^{2\beta+1}\mathrm{d}u\Big) \lesssim \frac{1}{c}, \quad (4.27)
\end{aligned}$$

which shows $\mathcal{E}_h = \mathcal{O}_P(m^{-1/2}h^{-\beta-1}\vee 1)$. □

Proof of Proposition 4.5

The following lemma establishes a bound for the bias term of the estimator for the distribution function.

Lemma 4.14. *Let Assumption 4.A hold with $\ell = \langle\alpha\rangle + 1$, $\alpha > 0$ and $f(\bullet + q_\tau) \in C^\alpha([-\zeta,\zeta], R)$. Then we have*

$$\sup_{f(\bullet+q_\tau)\in C^\alpha([-\zeta,\zeta],R)} \Big| \int_{-\infty}^{q_\tau} K_h * f(x)\mathrm{d}x - \int_{-\infty}^{q_\tau} f(x)\mathrm{d}x\Big| \leqslant Dh^{\alpha+1},$$

where $D = (R/(\langle\alpha\rangle+1)! + 2\zeta^{-\alpha-1})\|K(x)x^{\alpha+1}\|_{L^1}$.

Proof. Let $F(x) := \int_{-\infty}^{x} f(y)\mathrm{d}y$. Fubini's theorem yields

$$\int_{-\infty}^{q_\tau} K_h * f(x)\mathrm{d}x = \int_{-\infty}^{\infty} K_h(x)F(q_\tau - x)\mathrm{d}x, \tag{4.28}$$

where $K_h(x) := h^{-1}K(x/h), x \in \mathbb{R}$. Therefore, the bias depends only locally on f. Note that $F(\bullet + q_\tau) \in C^{\alpha+1}([-\zeta, \zeta])$ by assumption. A Taylor expansion of F around q_τ yields for $|hz| < \zeta$

$$F(q_\tau - hz) - F(q_\tau) = -hzF'(q_\tau) + \cdots + (-hz)^{\langle\alpha\rangle+1}\frac{F^{(\langle\alpha\rangle+1)}(q_\tau - \kappa hz)}{(\langle\alpha\rangle + 1)!},$$

where $0 \leqslant \kappa \leqslant 1$. Using the fact that $\int x^k K(x)\mathrm{d}x = 0$ for $k = 1, ..., \langle\alpha\rangle + 1$ and the properties of the class, we obtain

$$\begin{aligned}
&\Big|\int_{-\infty}^{q_\tau} (K_h * f(x) - f(x))\mathrm{d}x\Big| = \Big|\int_{-\infty}^{\infty} K(z)(F(q_\tau - hz) - F(q_\tau))\mathrm{d}z\Big| \\
=&\Big|\int_{|z|<\zeta/h} K(z)(-hz)^{\langle\alpha\rangle+1}\frac{F^{(\langle\alpha\rangle+1)}(q_\tau - \kappa hz) - F^{(\langle\alpha\rangle+1)}(q_\tau)}{(\langle\alpha\rangle + 1)!}\mathrm{d}z\Big| \\
&+ \int_{|z|\geqslant\zeta/h} |K(z)||F(q_\tau - hz) - F(q_\tau)|\mathrm{d}z \\
\leqslant&\frac{h^{\langle\alpha\rangle+1}R}{(\langle\alpha\rangle + 1)!}\int_{-\infty}^{\infty} |K(z)||z|^{\langle\alpha\rangle+1}|\kappa hz|^{\alpha+1-(\langle\alpha\rangle+1)}\mathrm{d}z + 2\int_{|z|\geqslant\zeta/h} |K_h(z)|\mathrm{d}z \\
\leqslant&\Big(\frac{h^{\alpha+1}R}{(\langle\alpha\rangle + 1)!} + 2\Big(\frac{h}{\zeta}\Big)^{\alpha+1}\Big)\int_{-\infty}^{\infty} |K(z)||z|^{\alpha+1}\mathrm{d}z,
\end{aligned}$$

and the statement follows. □

Proof of Proposition 4.5. Uniformly over $f \in \mathcal{C}^\alpha(R, r, \zeta)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$ we will show for any h such that $(n \wedge m)h^{2\beta+1} \to \infty$

$$\begin{aligned}
&\Big|\int_{-\infty}^{q_\tau} (\widetilde{f}_h(x) - f(x))\mathrm{d}x\Big| \\
&\quad= \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}\Big(h^{\alpha+1} + ((n \wedge m)(h^{2\beta-1} \wedge 1))^{-1/2} + ((n \wedge m)(mh^{2\beta+2} \wedge 1))^{-1/2}\Big).
\end{aligned}$$

The third term on the right-hand side is of smaller or of the same order than the second one if and only if $(mh^{1\wedge2\beta+2})^{-1} \lesssim 1$. Hence, when $\alpha \geqslant 1/2$ the asymptotically optimal choice $h = (n \wedge m)^{-1/(2\alpha+2(\beta\vee1/2)+1)}$ yields

$$\Big|\int_{-\infty}^{q_\tau} (\widetilde{f}_h(x) - f(x))\mathrm{d}x\Big| = \mathcal{O}_P\Big((n \wedge m)^{-(\alpha+1)/(2\alpha+2\beta+1)} \vee (n \wedge m)^{-1/2}\Big).$$

Step 1: As usual we decompose the error into a deterministic error term and a stochastic error term, writting $\varphi_X = \mathcal{F}f$,

$$\begin{aligned}
\Big|\int_{-\infty}^{q_\tau} \widetilde{f}_h(x) - f(x)\mathrm{d}x\Big| \leqslant&\Big|\int_{-\infty}^{q_\tau} K_h * f(x) - f(x)\mathrm{d}x\Big| \\
&+ \Big|\int_{-\infty}^{q_\tau} \mathcal{F}^{-1}\Big[\frac{\varphi_n(u)\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)} - \varphi_K(hu)\varphi_X(u)\Big](x)\mathrm{d}x\Big|.
\end{aligned}$$

The bias is of order $\mathcal{O}(h^{\alpha+1})$ by Lemma 4.14. As discussed above, we decompose the stochastic error into a singular part and a continuous one using a smooth truncation function. Let $a_c \in C^\infty(\mathbb{R})$ satisfy $a_c(x) = 1$ for $x \leqslant -1$ and $a_c(x) = 0$ for $x \geqslant 0$ and define $a_s(x) := \mathbb{1}_{(-\infty,0]}(x) - a_c(x)$. Then

$$\begin{aligned}
&\int_{-\infty}^{q_\tau} \mathcal{F}^{-1}\Big[\varphi_K(hu)\Big(\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)} - \varphi_X(u)\Big)\Big](x)\mathrm{d}x\\
&= \int a_s(x)\,\mathcal{F}^{-1}\Big[\varphi_K(hu)\Big(\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)} - \varphi_X(u)\Big)\Big](x+q_\tau)\mathrm{d}x\\
&\quad + \int a_c(x)\,\mathcal{F}^{-1}\Big[\varphi_K(hu)\Big(\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)} - \varphi_X(u)\Big)\Big](x+q_\tau)\mathrm{d}x =: T_s + T_c. \qquad (4.29)
\end{aligned}$$

The singular term T_s will be treated in the next step while we bound the continuous, but not integrable term T_c in Step 3.

Step 2: Lemma 4.13 shows that the probability of the complement $B_\varepsilon(h)^c$ of $B_\varepsilon(h)$ from (4.6) converges to zero. We obtain for any $c > 0$ with Markov's inequality

$$\begin{aligned}
&P\Big(|T_s| > \frac{c}{\sqrt{(n\wedge m)(h^{2\beta-1}\vee 1)}}\Big)\\
&\leqslant P\Big(B_\varepsilon(h) \cap \Big\{|T_s| > \frac{c}{\sqrt{(n\wedge m)(h^{2\beta-1}\vee 1)}}\Big\}\Big) + P\Big(B_\varepsilon(h)^c\Big)\\
&\leqslant \frac{1}{c}\sqrt{(n\wedge m)(h^{2\beta-1}\vee 1)}\,\mathbb{E}\Big[|T_s|\mathbb{1}_{B_\varepsilon(h)}\Big] + o(1).
\end{aligned}$$

To bound $\mathbb{E}[|T_s|\mathbb{1}_{B_\varepsilon(h)}]$, we first note by Plancherel's identity

$$\begin{aligned}
T_s =& \frac{1}{2\pi}\int \mathcal{F} a_s(u)e^{-iuq_\tau}\varphi_K(hu)\Big(\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)} - \varphi_X(u)\Big)\mathrm{d}u\\
=& \frac{1}{2\pi}\int \mathcal{F} a_s(u)e^{-iuq_\tau}\varphi_K(hu)\Big(\frac{\varphi_n(u)}{\varphi_\varepsilon(u)} - \varphi_X(u)\Big)\mathrm{d}u\\
&+ \frac{1}{2\pi}\int \mathcal{F} a_s(u)e^{-iuq_\tau}\frac{\varphi_K(hu)\varphi_n(u)}{\varphi_\varepsilon(u)}\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)} - 1\Big)\mathrm{d}u =: \frac{1}{2\pi}(T_{s,x} + T_{s,\varepsilon}). \qquad (4.30)
\end{aligned}$$

The first term $T_{s,x}$ corresponds to the error due to the unknown density f while $T_{s,\varepsilon}$ is dominated by the error of the estimator $\varphi_{\varepsilon,m}$. Since a_s is of bounded variation and has compact support, there is a constant $A_s \in (0,\infty)$ such that $|\mathcal{F} a_s(u)| \leqslant A_s(1+|u|)^{-1}$. Plancherel's identity yields

$$\begin{aligned}
\mathrm{Var}(T_{s,x}) = \mathbb{E}[|T_{s,x}|^2] \leqslant& \frac{1}{n}\mathbb{E}\Big[\Big|\int \mathcal{F} a_s(u)e^{-iuq_\tau}\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}e^{iuY_1}\mathrm{d}u\Big|^2\Big] \qquad (4.31)\\
\leqslant& \frac{2\pi}{n}\|f_Y\|_\infty\Big\|\mathcal{F}^{-1}\Big[\mathcal{F} a_s(u)\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}\Big]\Big\|_{L^2}^2\\
\leqslant& \frac{1}{n}\|K\|_{L^1}^2\|f_Y\|_\infty\int_{-1/h}^{1/h}\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\\
\leqslant& \frac{1}{n}\|K\|_{L^1}^2 A_s^2\|f_Y\|_\infty\int_{-1/h}^{1/h}\frac{1}{(1+|u|)^2|\varphi_\varepsilon(u)|^2}\mathrm{d}u.
\end{aligned}$$

Using the assumption $\|f\|_\infty < R$ and $f_\varepsilon \in \mathcal{D}^\beta(R,\gamma)$, we get

$$\mathbb{E}[|T_{s,x}|^2] \lesssim \frac{1}{n}\int_{-1/h}^{1/h}(1+|u|)^{2\beta-2}\mathrm{d}u \lesssim \frac{1}{nh^{2\beta-1}} \vee \frac{1}{n}. \tag{4.32}$$

To bound $T_{s,\varepsilon}$, we will use the following version of a lemma by Neumann (1997): By the definition (4.6) of $B_\varepsilon(h)$ and applying (4.23) it holds

$$\begin{aligned}\mathbb{E}\Big[\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathbb{1}_{B_\varepsilon(h)}\Big] &\leqslant 2\,\mathbb{E}\Big[\frac{|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^2}{|\varphi_\varepsilon(u)|^2}\Big]+2\,\mathbb{E}\Big[\frac{|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^4}{|\varphi_\varepsilon(u)\varphi_{\varepsilon,m}(u)|^2}\mathbb{1}_{B_\varepsilon(h)}\Big]\\ &\leqslant \frac{2\,\mathbb{E}[|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^2]}{|\varphi_\varepsilon(u)|^2}+\frac{2m\,\mathbb{E}[|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^4]}{|\varphi_\varepsilon(u)|^2}\\ &\leqslant \frac{18}{m|\varphi_\varepsilon(u)|^2}.\end{aligned} \tag{4.33}$$

We estimate with the Cauchy–Schwarz inequality

$$\begin{aligned}T_{s,\varepsilon}^2 &\leqslant \|K\|_{L^1}^2\int_{-1/h}^{1/h}\Big|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\Big|^2\mathrm{d}u\int_{-1/h}^{1/h}|\mathcal{F}a_s(u)|^2\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathrm{d}u\\ &\leqslant 2\|K\|_{L^1}^2\Big(\|\varphi_X\|_{L^2}^2+\int_{-1/h}^{1/h}\frac{|\varphi_n(u)-\varphi_Y(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)\\ &\qquad\times\int_{-1/h}^{1/h}|\mathcal{F}a_s(u)|^2\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathrm{d}u.\end{aligned}$$

Applying again the Cauchy–Schwarz inequality, Fubini's theorem, the decay of $\mathcal{F}a_s$ and (4.33), we obtain

$$\begin{aligned}\mathbb{E}[|T_{s,\varepsilon}|\mathbb{1}_{B_\varepsilon(h)}] \leqslant&\sqrt{2}\|K\|_{L^1}\Big(\|\varphi_X\|_{L^2}^2+\int_{-1/h}^{1/h}\frac{\mathbb{E}[|\varphi_n(u)-\varphi_Y(u)|^2]}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)^{1/2}\\ &\times\Big(\int_{-1/h}^{1/h}\frac{A_s^2}{(1+|u|)^2}\,\mathbb{E}\Big[\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathbb{1}_{B_\varepsilon(h)}\Big]\mathrm{d}u\Big)^{1/2}\\ \leqslant&\frac{\sqrt{36}\|K\|_{L^1}A_s}{\sqrt{m}}\Big(\|\varphi_X\|_{L^2}^2+\int_{-1/h}^{1/h}\frac{\mathrm{d}u}{n|\varphi_\varepsilon(u)|^2}\Big)^{1/2}\\ &\times\Big(\int_{-1/h}^{1/h}\frac{\mathrm{d}u}{(1+|u|)^2|\varphi_\varepsilon(u)|^2}\Big)^{1/2}.\end{aligned} \tag{4.34}$$

The assumptions $\|f\|_\infty \lesssim 1, |\varphi_\varepsilon(u)| \lesssim (1+|u|)^{-\beta}$ and $n^{-1}h^{-2\beta-1}\to 0$ for the optimal $h=h^*$ yield

$$\mathbb{E}[|T_{s,\varepsilon}|\mathbb{1}_{B_\varepsilon(h)}] \lesssim \Big(1+\frac{1}{nh^{2\beta+1}}\Big)^{1/2}\Big(\frac{1}{\sqrt{m}h^{\beta-1/2}}\vee\frac{1}{\sqrt{m}}\Big) \lesssim \frac{1}{\sqrt{m}h^{\beta-1/2}}\vee\frac{1}{\sqrt{m}}.$$

Together with (4.32) and (4.30) this implies the optimal order

$$\mathbb{E}[|T_s|\mathbb{1}_{B_\varepsilon(h)}] \lesssim \Big((n\wedge m)(h^{2\beta-1}\wedge 1)\Big)^{-1/2}.$$

Step 3: The empirical measures of (Y_j) and (ε_k) are given by $\mu_{Y,n} := \frac{1}{n}\sum_{j=1}^n \delta_{Y_j}$ and $\mu_{\varepsilon,m} := \frac{1}{m}\sum_{k=1}^m \delta_{\varepsilon_k}$, respectively, with Dirac measure δ_x in $x \in \mathbb{R}$. We can write

$$\begin{aligned}
T_c &= \int a_c(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big(\varphi_n(u) - \varphi_{\varepsilon,m}(u)\varphi_X(u)\Big)\Big](x+q_\tau)\mathrm{d}x \\
&= \mathcal{F}^{-1}\Big[\frac{\varphi_K(-hu)}{\varphi_{\varepsilon,m}(-u)}\Big(\varphi_n(-u) - \varphi_{\varepsilon,m}(-u)\varphi_X(-u)\Big)\Big] * a_c(-q_\tau) \\
&= \mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big] * \Big(\mu_{Y,n} * a_c(-\bullet) - \mu_{\varepsilon,m} * f * a_c(-\bullet)\Big)(q_\tau).
\end{aligned}$$

Applying Lemma 4.2, we obtain on $B_\varepsilon(h)$ for any integer $s > \beta$

$$\begin{aligned}
|T_c| &\leqslant \Big\|\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_{\varepsilon,m}(u)}\Big] * \Big(\mu_{Y,n} * a_c(-\bullet) - \mu_{\varepsilon,m} * f * a_c(-\bullet)\Big)\Big\|_\infty \\
&\leqslant \mathcal{E}_h\Big\|\mu_{Y,n} * a_c(-\bullet) - \mu_{\varepsilon,m} * f * a_c(-\bullet)\Big\|_{C^s} \\
&= \mathcal{E}_h\sum_{l=0}^s \|\mu_{Y,n} * a_c^{(l)}(-\bullet) - \mu_{\varepsilon,m} * f * a_c^{(l)}(-\bullet)\|_\infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&P\Big(B_\varepsilon(h) \cap \Big\{|T_c| > \frac{c}{\sqrt{(n\wedge m)}(\sqrt{m}h^{\beta+1}\wedge 1)}\Big\}\Big) \\
&\leqslant P\Big(B_\varepsilon(h) \cap \Big\{\mathcal{E}_h > \Big(\frac{c}{mh^{2\beta+2}\wedge 1}\Big)^{1/2}\Big\}\Big) \\
&\quad + P\Big(\sum_{l=0}^s \|\mu_{Y,n} * a_c^{(l)} - \mu_{\varepsilon,m} * f * a_c^{(l)}\|_\infty > \Big(\frac{c}{n\wedge m}\Big)^{1/2}\Big) =: P_1 + P_2.
\end{aligned}$$

By Lemma 4.2, more precisely estimate (4.27), the first probability is of the order $1/c$. To bound P_2, it suffices to show $\|\mu_{Y,n} * a_c^{(l)} - \mu_{\varepsilon,m} * f * a_c^{(l)}\|_\infty = \mathcal{O}_P((n\wedge m)^{-1/2})$ for all $l = 0, \dots, s$. Denoting the density of Y_j as $f_Y = f * f_\varepsilon$, we decompose

$$\begin{aligned}
&\|\mu_{Y,n} * (a_c^{(l)}(-\bullet)) - \mu_{\varepsilon,m} * f * (a_c^{(l)}(-\bullet))\|_\infty \\
&\quad\leqslant \|\mu_{Y,n} * (a_c^{(l)}(-\bullet)) - f_Y * (a_c^{(l)}(-\bullet))\|_\infty \\
&\qquad + \|f_\varepsilon * (f * (a_c^{(l)}(-\bullet))) - \mu_{\varepsilon,m} * (f * (a_c^{(l)}(-\bullet)))\|_\infty \\
&\quad\leqslant \Big\|\int a_c^{(l)}(y-\bullet)\mu_{Y,n}(\mathrm{d}y) - \mathbb{E}[a_c^{(l)}(Y_1 - \bullet)]\Big\|_\infty \\
&\qquad + \Big\|\mathbb{E}[(f * a_c^{(l)})(\varepsilon_1 - \bullet)] - \int (f * a_c^{(l)})(z-\bullet)\mu_{\varepsilon,m}(\mathrm{d}z)\Big\|_\infty
\end{aligned}$$

By construction all $a_c^{(l)}, l \geqslant 1$, have compact support and are bounded. Hence, $\|a_c^{(l)}\|_{L^1} < \infty$, $\|(a_c * f)^{(l)}\|_{L^1} \leqslant \|a_c^{(l)}\|_{L^1}\|f\|_{L^1} < \infty$ and thus $a_c^{(l)}(\bullet - t)$ and $a_c^{(l)} * f(\bullet - t)$, $l \geq 0$, are of bounded variation for all $t \in \mathbb{R}$. Since the set of functions with bounded variation is a Donsker class (cf. Theorem 2.1 by Dudley (1992)), the two terms in the previous display converge in probability to a tight limit with $\sqrt{n}$-rate and $\sqrt{m}$-rate, respectively. Consequently,

$$\sqrt{n\wedge m}\|\mu_{Y,n} * (a_c^{(l)}(-\bullet)) - \mu_{\varepsilon,m} * f * (a_c^{(l)}(-\bullet))\|_\infty = \mathcal{O}_P(1)$$

for all $\ell = 0, \dots, s$ and P_2 is arbitrary small for c large. $\square$

For the adaptive estimator we will later need the following uniform version of Proposition 4.5.

Corollary 4.15. *Suppose Assumption 4.A holds with $l = \langle\alpha\rangle + 1$ and let $\mathcal{B}$ be a finite set of bandwidths with $h_1 = \min \mathcal{B}$ such that $mh_1^{2\beta_1+1} \to \infty$. For critical values $(\delta_h)_{h\in\mathcal{B}}$ satisfying $\delta_h > 3Dh^{\alpha+1}$ and for any sequence $(x_n)_n$ with $x_n \to \infty$ arbitrarily slowly we obtain uniformly in $\mathcal{C}^\alpha(R, r, \zeta)$ and $\mathcal{D}^\beta(R, \gamma)$*

$$P\Big(\exists h \in \mathcal{B} : \Big|\int_{-\infty}^{q_\tau} (\widetilde{f}_h(x) - f(x))\mathrm{d}x\Big| > \delta_h\Big)$$
$$= \mathcal{O}\Big(\sum_{h\in\mathcal{B}} \Big(\frac{1}{\delta_h}((n\wedge m)(h^{2\beta-1}\wedge 1))^{-1/2} + \frac{1}{\delta_h^2}\frac{x_n}{(n\wedge m)(mh^{2\beta+2}\wedge 1)}\Big)\Big) + o(1).$$

In particular, if $|\mathcal{B}| \lesssim \log n, \max_{h\in\mathcal{B}} h \to 0$ and $\min_{h\in\mathcal{B}}(n\wedge m)h^{2\beta+1} \to \infty$, then

$$\sup_{h\in\mathcal{B}}\Big|\int_{-\infty}^{q_\tau} \widetilde{f}_h(x) - f(x)\mathrm{d}x\Big| \xrightarrow{P} 0.$$

Proof. With the notation of the proof of Proposition 4.5 and applying Lemma 4.14, we obtain

$$\Big|\int_{-\infty}^{q_\tau} (\widetilde{f}_h(x) - f(x))\mathrm{d}x\Big| \leqslant \Big|\int_{-\infty}^{q_\tau} (K_h * f(x) - f(x))\mathrm{d}x\Big| + |T_s| + |T_c| \leqslant Dh^{\alpha+1} + |T_s| + |T_c|,$$

where T_s and T_c are the stochastic errors of the singular part and of the continuous part, respectively, as defined in (4.29). Since both terms depend on h let us write $T_s(h)$ and $T_c(h)$. By definition $h_1 \leqslant h$ implies $B_\varepsilon(h_1) \subseteq B_\varepsilon(h)$. Then, Step 2 in the previous proof shows

$$\begin{aligned}
P(\exists h \in \mathcal{B} : T_s > \delta_h/3) &\leqslant \Big(\sum_{h\in\mathcal{B}} P(\{T_s(h) > \delta_h/3\} \cap B_\varepsilon(h_1))\Big) + o(1)\\
&\leqslant \Big(\sum_{h\in\mathcal{B}} \delta_h^{-1}\,\mathbb{E}[|T_s(h)|\mathbb{1}_{B_\varepsilon(h_1)}]\Big) + o(1)\\
&\lesssim \Big(\sum_{h\in\mathcal{B}} \delta_h^{-1}((n\wedge m)(h^{2\beta-1}\wedge 1))^{-1/2}\Big) + o(1).
\end{aligned}$$

Following Step 3 in the previous proof, we obtain with the random operator norm $\mathcal{E}_h$, for some integer $s > \beta$ and for a diverging sequence $(x_{(n\wedge m)})$

$$\begin{aligned}
&P(\exists h \in \mathcal{B} : T_c > \delta_h/3)\\
\leqslant &P(\{\exists h \in \mathcal{B} : \mathcal{E}_h > \delta_h(n\wedge m)^{1/2}/(3(x_{(n\wedge m)})^{1/2})\} \cap B_\varepsilon(h_1)) + P(B_\varepsilon(h_1)^c)\\
&+ P\Big(\Big\{\sum_{l=0}^{s} \|\mu_{Y,n} * a_c^{(l)} - \mu_{\varepsilon,m} * f * a_c^{(l)}\|_\infty > \Big(\frac{x_{(n\wedge m)}}{(n\wedge m)}\Big)^{1/2}\Big\}\Big)\\
\leqslant &\Big(\sum_{h\in\mathcal{B}} P(\{\mathcal{E}_h > \delta_h(n\wedge m)^{1/2}/(3(x_{(n\wedge m)})^{1/2})\} \cap B_\varepsilon(h_1))\Big) + o(1)\\
\lesssim &\Big(\sum_{h\in\mathcal{B}} \frac{x_{(n\wedge m)}}{\delta_h^2(n\wedge m)(mh^{2\beta+2}\wedge 1)}\Big) + o(1),
\end{aligned}$$

where we have used (4.27) in the last estimate. $\square$

Proof of Proposition 4.6

Without loss of generality we set $q_\tau = 0$. Recall definition (4.12) of the pseudo-estimator $\widehat{f}_h$ which knows the error distribution. We estimate

$$\begin{aligned}\sup_{x\in(-\zeta,\zeta)} |\widetilde{f}_h(x) - f(x)| &\leqslant \sup_{x\in(-\zeta,\zeta)} |\widehat{f}_h(x) - f(x)| + \|\widetilde{f}_h - \widehat{f}_h\|_\infty \\ &\leqslant \sup_{x\in(-\zeta,\zeta)} |\widehat{f}_h(x) - f(x)| + \Big\|\frac{\varphi_K(hu)\varphi_n(u)}{\varphi_\varepsilon(u)}\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)} - 1\Big)\Big\|_{L^1}.\end{aligned}$$

The analysis of the first term is very classical. However, we are not aware of any reference in the given setup. Both terms will be treated separately in the following two steps. All estimates will be uniform in $f \in \mathcal{C}^\alpha(R, r, \zeta)$ and $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$.

Step 1: Let $h \in (0, 1)$. We will show that there are constants $d, D > 0$ such that for any $t > d(h^\alpha + (nh^{2\beta+1})^{-2})$

$$P\Big(\sup_{x\in(-\zeta,\zeta)} |\widehat{f}_h(x) - f(x)| > t\Big) \leqslant 2\exp\Big(2\log n - Dnh^{(2\beta+1)}(t \wedge t^2)\Big). \tag{4.35}$$

Then the result follows by choosing $t \sim h^\alpha + \big(\frac{\log n}{nh^{2\beta+1}}\big)^{1/2}$. Let us define $x_k := -\zeta + kn^{-2}$ for $k = 1, \dots, \lfloor 2\zeta n^2\rfloor =: M$ as well as

$$\begin{aligned}\chi_j(x) :=& \mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}e^{iuY_j}\Big](x) - \mathbb{E}\Big[\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}e^{iuY_j}\Big](x)\Big] \\ =& K_h * \mathcal{F}^{-1}\Big[\mathbb{1}_{[-h^{-1},h^{-1}]}(u)\frac{e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x) - K_h * f(x), \quad x \in \mathbb{R}.\end{aligned}$$

Therefore, $\widehat{f}_h(x) - \mathbb{E}[\widehat{f}_h(x)] = \frac{1}{n}\sum_{j=1}^n \chi_j(x)$ and thus

$$\begin{aligned}&\sup_{|x|<\zeta} |\widehat{f}_h(x) - f(x)| \leqslant \sup_{|x|<\zeta} |\mathbb{E}[\widehat{f}_h(x)] - f(x)| + \sup_{|x|<\zeta} |\widehat{f}_h(x) - \mathbb{E}[\widehat{f}_h(x)]| \\ &\leqslant \sup_{|x|<\zeta} |\mathbb{E}[\widehat{f}_h(x)] - f(x)| + \sup_{|x|<\zeta}\min_{1\leqslant k\leqslant M}\Big|\frac{1}{n}\sum_{j=1}^n (\chi_j(x) - \chi_j(x_k))\Big| + \max_{1\leqslant k\leqslant M}\Big|\frac{1}{n}\sum_{j=1}^n \chi_j(x_k)\Big| \\ &=: B + V_1 + V_2.\end{aligned}$$

The bias term B can be bounded as in the classical density estimation setup (cf. also Fan, 1991b, Thm. 1 and 2), noting that the constant does not depend on $x \in (-\zeta, \zeta)$. Hence, $|B| \lesssim h^\alpha$. Using a continuity argument and the properties of $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$, the term V_1 can be bounded by

$$\begin{aligned}|V_1| &\leqslant \frac{1}{n^2}\Big\|\frac{1}{n}\sum_{j=1}^n \chi_j'\Big\|_\infty = \frac{1}{n^3}\Big\|\sum_{j=1}^n (K_h') * \Big(\mathcal{F}^{-1}\Big[\mathbb{1}_{[-h^{-1},h^{-1}]}(u)\frac{e^{iuY_j}}{\varphi_\varepsilon(u)}\Big] - f\Big)\Big\|_\infty \\ &\leqslant \frac{1}{n^2h}\|K'\|_{L^1}\big(\|\mathbb{1}_{[-h^{-1},h^{-1}]}\varphi_\varepsilon^{-1}\|_{L^1} + \|f\|_\infty\big) \lesssim n^{-2}h^{-(\beta+2)} \lesssim (nh^{2\beta+1})^{-2}.\end{aligned}$$

Therefore, $|B+V_1| \leqslant D_1(h^\alpha + (nh^{2\beta+1})^{-2})$ for some constant $D_1 > 0$. We obtain for all $t > d(h^\alpha + (nh^{2\beta+1})^{-2})$ with $d := 2D_1$

$$P\Big(\sup_{|x|<\zeta} |\widehat{f}_h(x) - f(x)| > t\Big) \leqslant P\Big(\max_{k=1,\dots,M} \Big|\frac{1}{n}\sum_{j=1}^n \chi_j(x_k)\Big| > \frac{t}{2}\Big)$$
$$\leqslant \sum_{k=1}^M P\Big(\Big|\frac{1}{n}\sum_{j=1}^n \chi_j(x_k)\Big| > \frac{t}{2}\Big).$$

Finally, we will apply Bernstein's inequality. To this end we estimate

$$\max_{j,k} |\chi_j(x_k)| \leqslant 2\|K_h\|_{L^1} \|\mathbb{1}_{[-h^{-1},h^{-1}]}\varphi_\varepsilon^{-1}\|_{L^1} \leqslant D_2 h^{-(\beta+1)},$$

with some constant $D_2 > 0$. Using Plancherel's identity, the variance can be estimated by

$$\begin{aligned}\operatorname{Var}(\chi_j(x_k)) &= \mathbb{E}\Big[\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)} e^{iuY_j}\Big]^2(x_k)\Big] - (K_h * f)^2(x_k)\\ &\leqslant \frac{1}{2\pi}\|f\|_\infty \Big\|\frac{\varphi_K(-hu)}{\varphi_\varepsilon(-u)}\Big\|_{L^2}^2 \lesssim D_3 h^{-(2\beta+1)},\end{aligned}$$

for some $D_3 > 0$. Then Bernstein's inequality yields

$$\begin{aligned}&P\Big(\sup_{x\in(-\zeta,\zeta)} |\widehat{f}_h(x) - f(x)| > t\Big) \leqslant \sum_{k=1}^M P\Big(|\sum_{j=1}^n \chi_j(x_k)| > nt/2\Big)\\ \leqslant& 2\exp\Big(\log M - \frac{nh^{(2\beta+1)}t^2}{8(D_3 + D_2 t/3)}\Big) \leqslant 2\exp\Big(2\log n - Dnh^{(2\beta+1)}(t \wedge t^2)\Big),\end{aligned}$$

with some constant $D > 0$.

Step 2: By the Cauchy–Schwarz inequality we have

$$\begin{aligned}&\mathbb{E}\Big[\Big\|\frac{\varphi_K(hu)\varphi_n(u)}{\varphi_\varepsilon(u)}\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)} - 1\Big)\Big\|_{L^1} \mathbb{1}_{B_\varepsilon(h)}\Big]\\ \lesssim&\Big(\mathbb{E}\Big[\Big\|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\mathbb{1}_{[-1/h,1/h]}(u)\Big\|_{L^2}^2\Big]\,\mathbb{E}\Big[\Big\|\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)} - 1\Big)\mathbb{1}_{[-1/h,1/h]}(u)\Big\|_{L^2}^2 \mathbb{1}_{B_\varepsilon(h)}\Big]\Big)^{1/2}\\ \leqslant&\Big(\|\varphi_X\|_{L^2} + \Big(\int_{-1/h}^{1/h} \frac{\mathbb{E}[|\varphi_n(u) - \varphi_Y(u)|^2]}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)^{1/2}\Big)\\ &\qquad\times\Big(\int_{-1/h}^{1/h} \mathbb{E}\Big[\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)} - 1\Big|^2 \mathbb{1}_{B_\varepsilon(h)}\Big]\mathrm{d}u\Big)^{1/2}\\ \lesssim&\Big(\|\varphi_X\|_{L^2} + \Big(\frac{1}{nh^{2\beta+1}}\Big)^{1/2}\Big)\Big(\frac{1}{mh^{2\beta+1}}\Big)^{1/2},\end{aligned}$$

where we have used (4.33) for the last step. Therefore, the additional error due to the unknown error distribution satisfies for any $\delta > 0$ by Markov's inequality and by

Lemma 4.13

$$
\begin{aligned}
P\Big(\Big\|\frac{\varphi_K(hu)\varphi_n(u)}{\varphi_\varepsilon(u)}\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big)\Big\|_{L^1}>\delta\Big)\\
&\leqslant \frac{1}{\delta}\mathbb{E}\Big[\Big\|\frac{\varphi_K(hu)\varphi_n(u)}{\varphi_\varepsilon(u)}\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big)\Big\|_{L^1}\mathbb{1}_{B_\varepsilon(h)}\Big]+P\Big(\inf_{|u|\leqslant 1/h}|\varphi_{\varepsilon,m}(u)|<m^{-1/2}\Big)\\
&\lesssim \frac{1}{\delta}\Big(\frac{1}{mh^{2\beta+1}}\Big)^{1/2}+o(1). \qquad (4.36)
\end{aligned}
$$

and thus $\|\widetilde{f}_h-\widehat{f}_h\|_\infty=\mathcal{O}_P((mh^{2\beta+1})^{-1/2})$. Note that the second term does not depend on δ and thus $o(1)$ is sufficient. □

Proof of Theorems 4.7 and 4.8

We start with a lemma that establishes consistency of the quantile estimator and then prove the theorems. To apply this lemma also for the adaptive result, we prove convergence uniformly over a set of bandwidths.

Lemma 4.16. *Grant Assumption 4.A with $\ell=1$. Let $\mathcal{B}$ be a set of bandwidths satisfying $|\mathcal{B}|\lesssim\log n,\max\mathcal{B}\to 0$ and $\min_{h\in\mathcal{B}}(\log n)^2/((n\wedge m)h^{2\beta+1})\to 0$. Then*

$$\sup_{f\in\mathcal{C}^\alpha(R,r,\zeta,U_n)}\sup_{f_\varepsilon\in\mathcal{D}^\beta(R,\gamma)}P\big(\sup_{h\in\mathcal{B}}|\widetilde{q}_{\tau,h}-q_\tau|>\delta\big)\to 0\quad\textit{for all }\delta>0.$$

Proof. We follow the general strategy of the proof of Theorem 5.7 by van der Vaart (1998) in the classical M-estimation setting. Recall the definition of $\widetilde{M}_h$ given in (4.5) and its deterministic counterpart $M(\eta)=\int_{-\infty}^{\eta}f(x)\mathrm{d}x-\tau$. To this end, we first show that f satisfies the uniqueness condition, for $\delta_0:=(\frac{r}{2R})^{1\vee\alpha^{-1}}\wedge\zeta>0$,

$$\inf_{\eta:|\eta-q_\tau|\geqslant\delta}|M(\eta)|>\frac{r}{2}(\delta\wedge\delta_0)\quad\text{for any }\delta>0. \qquad (4.37)$$

By the Hölder regularity $M'(\eta)=f(\eta)\geqslant f(q_\tau)-|f(q_\tau)-f(\eta)|\geqslant r-R|q_\tau-\eta|^{1\wedge\alpha}\geqslant r/2$ for $|q_\tau-\eta|\leqslant\delta_0$. Recall that q_τ is given by the root of M and that M is increasing. Hence, we obtain

$$\inf_{\eta:|\eta-q_\tau|\geqslant\delta}|M(\eta)|=\inf_{\eta\in\{-\delta,\delta\}}\big|M(q_\tau-\eta)-M(q_\tau)\big|\geqslant(\delta\wedge\delta_0)\inf_{\eta:|\eta-q_\tau|\geqslant\delta}M'(\eta)\geqslant\frac{r}{2}(\delta\wedge\delta_0).$$

which gives (4.38).

Let us now construct an event A with $P(A)\to 1$ such that

$$\sup_{h\in\mathcal{B}}\widetilde{M}_h(\widetilde{q}_{\tau,h})=0\quad\text{on }A. \qquad (4.38)$$

Using $M(q_\tau)=0$ and (4.37), we conclude for $\delta\in(0,\delta_0)$

$$M(q_\tau-\delta)\leqslant-\frac{\delta r}{2}<0<\frac{\delta r}{2}\leqslant M(q_\tau+\delta)$$

The equality $|\widetilde{M}_h(\eta)-M(\eta)| = |\int_{-\infty}^{\eta}(\widetilde{f}_h - f)(x)\mathrm{d}x|$ and Corollary 4.15 imply uniformly over $h\in\mathcal{B}$ and $\eta\in\{q_\tau-\delta, q_\tau+\delta\}$

$$A := \{\forall h\in\mathcal{B}, \forall\eta\in\{q_\tau-\delta,q_\tau+\delta\} : |\widetilde{M}_h(\eta)-M(\eta)|\leqslant \frac{\delta r}{4}\} \quad\text{satisfies}\quad P(A)\to 1. \tag{4.39}$$

We conclude on A that $\widetilde{M}_h(q_\tau-\delta)<0<\widetilde{M}_h(q_\tau+\delta)$ for all $h\in\mathcal{B}$. Due to $\|\widetilde{f}_h\|_\infty \leqslant \|\varphi_K(b\bullet)/\varphi_\varepsilon\|_{L^1}$, $\widetilde{M}_h$ is continuous and thus there is an intermediate point $\xi_h\in(q_\tau-\delta, q_\tau+\delta)$ such that $\widetilde{M}_h(\xi_h)=0$. Since $\widetilde{q}_{\tau,h}$ minimizes $\widetilde{M}_h$ on the interval $[-U_n,U_n]$ which contains ξ_h for δ sufficiently small, we obtain (4.38).

Applying (4.37) and (4.38) yield

$$\begin{aligned}
P\big(\sup_{h\in\mathcal{B}}|\widetilde{q}_{\tau,h}-q_\tau|>\delta\big) &\leqslant P\big(\sup_{h\in\mathcal{B}}|M(\widetilde{q}_{\tau,h})|\geqslant \delta r/2\big)\\
&= P\big(\sup_{h\in\mathcal{B}}|M(\widetilde{q}_{\tau,h})-\widetilde{M}_h(\widetilde{q}_{\tau,h})|\geqslant\delta r/2\big)+P(A^c)\\
&\leqslant P\big(\sup_{h\in\mathcal{B}}\sup_{\eta\in[-U_n,U_n]}|M(\eta)-\widetilde{M}_h(\eta)|\geqslant \delta r/2\big)+o(1)\\
&= P\Big(\sup_{h\in\mathcal{B}}\sup_{\eta\in[-U_n,U_n]}\Big|\int_{-\infty}^{\eta}(\widetilde{f}_h(x)-f(x))\mathrm{d}x\Big|\geqslant\delta r/2\Big)+o(1).
\end{aligned} \tag{4.40}$$

Hence, it remains to show uniform consistency of $\int_{-\infty}^{\eta}\widetilde{f}_h(x)\mathrm{d}x$. Write

$$\begin{aligned}
\Big|\int_{-\infty}^{\eta}(\widetilde{f}_h(x)-f(x))\mathrm{d}x\Big| \leqslant&\Big|\int_{-\infty}^{\eta}(K_h*f(x)-f(x))\mathrm{d}x\Big|+\Big|\int_{-\infty}^{\eta}(\widetilde{f}_h(x)-K_h*f(x))\mathrm{d}x\Big|\\
=&|K_h*F(\eta)-F(\eta)|+\Big|\int_{-\infty}^{\eta}(\widetilde{f}_h(x)-K_h*f(x))\mathrm{d}x\Big|.
\end{aligned}$$

We have $|K_h*F(\eta)-F(\eta)| = |\int K_h(z)(F(\eta-z)-F(\eta))\mathrm{d}z|\leqslant h\|f\|_\infty\|zK(z)\|_{L^1}$ by the boundedness of f. Further note for $\eta\in[-U_n,U_n]$

$$\begin{aligned}
&\Big|\int_{-\infty}^{\eta}(\widetilde{f}_h(x)-K_h*f(x))\mathrm{d}x\Big|\\
\leqslant&\Big|\int_{-\infty}^{q_\tau}(\widetilde{f}_h(x)-K_h*f(x))\mathrm{d}x\Big|+\Big|\int_{q_\tau\wedge\eta}^{q_\tau\vee\eta}(\widetilde{f}_h(x)-K_h*f(x))\mathrm{d}x\Big|\\
\leqslant&\Big|\int_{-\infty}^{q_\tau}(\widetilde{f}_h(x)-K_h*f(x))\mathrm{d}x\Big|+\sqrt{2U_n}\Big(\int_{-\infty}^{\infty}(\widetilde{f}_h(x)-K_h*f(x))^2\mathrm{d}x\Big)^{1/2},
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality for the last step. Hence, together with (4.40) we obtain for all $\delta>6\|f\|_\infty\|zK(z)\|_{L^1}/r\sup_{h\in\mathcal{B}}h$

$$\begin{aligned}
P\big(\sup_{h\in\mathcal{B}}|\widetilde{q}_{\tau,h}-q_\tau|>\delta\big)\leqslant&P\Big(\sup_{h\in\mathcal{B}}\sup_{\eta\in[-U_n,U_n]}\Big|\int_{-\infty}^{\eta}(\widetilde{f}_h(x)-f(x))\mathrm{d}x\Big|\geqslant\delta r/2\Big)+o(1)\\
\leqslant&P\Big(\sup_{h\in\mathcal{B}}\Big|\int_{-\infty}^{q_\tau}(\widetilde{f}_h(x)-K_h*f(x))\mathrm{d}x\Big|\geqslant\frac{\delta r}{6}\Big)\\
&+P\Big(\sup_{h\in\mathcal{B}}\int(\widetilde{f}_h(x)-K_h*f(x))^2\mathrm{d}x\geqslant\frac{\delta^2r^2}{72U_n}\Big).
\end{aligned}$$

Corollary 4.15 shows under the conditions on $\mathcal{B}$ that

$$P\Big(\sup_{h\in\mathcal{B}}\Big|\int_{-\infty}^{q_\tau}\widetilde{f}_h(x)-K_h*f(x)\mathrm{d}x\Big|>\delta r/6\Big)\to 0$$

Hence, it remains to show

$$P\Big(\sup_{h\in\mathcal{B}}\int(\widetilde{f}_h(x)-K_h*f(x))^2\mathrm{d}x>\delta^2r^2/(72U_n)\Big)\to 0. \tag{4.41}$$

On the event $B_\varepsilon(h)$, (4.41) follows basically from the work of Neumann (1997). More precisely, Plancherel's equality, (4.33) and the Cauchy–Schwarz inequality yield for any $h\in\mathcal{B}$

$$\begin{aligned}
&\mathbb{E}\Big[\int(\widetilde{f}_h(x)-K_h*f(x))^2\mathrm{d}x\mathbb{1}_{B_\varepsilon(h)}\Big]\\
=&\frac{1}{2\pi}\int|\varphi_K(hu)|^2\,\mathbb{E}\Big[\Big|\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)}-\frac{\varphi_Y(u)}{\varphi_\varepsilon(u)}\Big|^2\mathbb{1}_{B_\varepsilon(h)}\Big]\mathrm{d}u\\
\lesssim&\int_{-1/h}^{1/h}\Big(\mathbb{E}\Big[\frac{|\varphi_n(u)-\varphi_Y(u)|^2}{|\varphi_{\varepsilon,m}(u)|^2}\mathbb{1}_{B_\varepsilon(h)}\Big]+|\varphi_Y(u)|^2\,\mathbb{E}\Big[\Big|\frac{1}{\varphi_{\varepsilon,m}(u)}-\frac{1}{\varphi_\varepsilon(u)}\Big|^2\mathbb{1}_{B_\varepsilon(h)}\Big]\Big)\mathrm{d}u\\
\lesssim&\int_{-1/h}^{1/h}\Big(\mathbb{E}\Big[\frac{|\varphi_n(u)-\varphi_Y(u)|^2}{|\varphi_\varepsilon(u)|^2}(1+m|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^2)\Big]+\frac{|\varphi_Y(u)|^2}{m|\varphi_\varepsilon(u)|^4}\Big)\mathrm{d}u\\
\leqslant&\int_{-1/h}^{1/h}\Big(\mathbb{E}\left[|\varphi_n(u)-\varphi_Y(u)|^4\right]\mathbb{E}\left[2+2m^2|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^4\right]\Big)^{1/2}\frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2}\\
&\qquad+\int_{-1/h}^{1/h}\frac{|\varphi_X(u)|^2}{m|\varphi_\varepsilon(u)|^2}\,\mathrm{d}u\\
\lesssim&\int_{-1/h}^{1/h}|\varphi_\varepsilon(u)|^{-2}(n^{-1}+m^{-1})\mathrm{d}u\lesssim\frac{1}{(n\wedge m)h^{2\beta+1}}.
\end{aligned}$$

Using $B_\varepsilon(\min\mathcal{B})\subseteq B_\varepsilon(h)$ and Lemma 4.13, (4.41) follows from Markov's inequality

$$\begin{aligned}
&P\Big(\sup_{h\in\mathcal{B}}\int(\widetilde{f}_h(x)-K_h*f(x))^2\mathrm{d}x>\delta^2r^2/(72U_n)\Big)\\
&\qquad\lesssim\frac{U_n}{\delta^2}\sum_{h\in\mathcal{B}}\mathbb{E}\Big[\int(\widetilde{f}_h(x)-K_h*f(x))^2\mathrm{d}x\mathbb{1}_{B_\varepsilon(\min\mathcal{B})}\Big]+P\big((B_\varepsilon(\min\mathcal{B}))^c\big)\\
&\qquad\lesssim\frac{(\log n)^2}{\delta^2(n\wedge m)h^{2\beta+1}}+o(1).
\end{aligned}$$

□

Proof of Theorem 4.7. A Taylor expansion yields

$$\begin{aligned}
\widetilde{q}_{\tau,h}-q_\tau=\frac{\widetilde{M}_h(\widetilde{q}_{\tau,h})-\widetilde{M}_h(q_\tau)}{\widetilde{M}_h'(q_\tau^*)}&=\frac{\widetilde{M}_h(\widetilde{q}_{\tau,h})-\int_{-\infty}^{q_\tau}\widetilde{f}_h(x)\mathrm{d}x+\tau}{\widetilde{f}_h(q_\tau^*)}\\
&=\frac{\widetilde{M}_h(\widetilde{q}_{\tau,h})-\int_{-\infty}^{q_\tau}(\widetilde{f}_h(x)-f(x))\mathrm{d}x}{\widetilde{f}_h(q_\tau^*)},
\end{aligned}\tag{4.42}$$

for some intermediate point q_τ^* between q_τ and $\widetilde{q}_{\tau,h}$. By Proposition 4.5 and (4.38), the numerator in the above display is of order $\mathcal{O}_P(n^{-(\alpha+1)/(2\alpha+2\beta+1)})$ for the optimal bandwidth h^*. For the denominator we will show $\widetilde{f}_h(q_\tau^*) = f(q_\tau)+o_p(1)$ which completes the proof. Since $f(\bullet + q_\tau) \in C^\alpha([-\zeta,\zeta],R)$, we obtain $|f(x+q_\tau) - f(q_\tau)| < t/2$ for all $|x| \leqslant (\frac{t}{2R})^{1\vee\alpha^{-1}} \wedge \zeta =: \delta$ for any $t > 0$. Therefore,

$$
\begin{aligned}
P(|\widetilde{f}_h(q_\tau^*) - f(q_\tau)| > t) &\leqslant P(\sup_{x\in[-\delta,\delta]} |\widetilde{f}_h(x+q_\tau) - f(q_\tau)| > t) + P(|\widetilde{q}_{\tau,h} - q_\tau| > \delta) \\
&\leqslant P(\sup_{x\in[-\delta,\delta]} |\widetilde{f}_h(x+q_\tau) - f(x+q_\tau)| > t/2) + P(|\widetilde{q}_{\tau,h} - q_\tau| > \delta). \qquad (4.43)
\end{aligned}
$$

Checking that the bandwidth satisfies $h \to 0$ and $\log(n)/(nh^{2\beta+1}) \to 0$ for $n \to \infty$, the first term on the right-hand side above converges to zero by the uniform consistency proved in Proposition 4.6. The second one vanishes asymptotically by Lemma 4.16. □

Proof of Theorem 4.8. Under the smoothness condition the interval (τ_1, τ_2) coincides with a bounded interval of quantiles (q_{τ_1}, q_{τ_2}). Noting that all our estimates are independent of the quantile, Theorem 4.8 can be proved along the same lines as Theorem 4.7 with only minor adaptation to $\sup_{\tau\in(\tau_1,\tau_2)}$ given a uniform version of Proposition 4.5: Uniformly over f in the class defined in the theorem and $f_\varepsilon \in \mathcal{D}^\beta(R,\gamma)$ for any h such that $(n \wedge m)h^{2\beta+1} \to \infty$ it holds

$$
\begin{aligned}
\sup_{\tau\in(\tau_1,\tau_2)} \Big| \int_{-\infty}^{q_\tau} (\widetilde{f}_h(x) - f(x))\mathrm{d}x \Big| =& \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}\Big(h^{\alpha+1} + (\tfrac{\log n}{n} \vee \tfrac{1}{m})^{1/2}(h^{-\beta+1/2} \vee 1) \\
&+ (\tfrac{1}{n} \vee \tfrac{1}{m})^{1/2}(m^{-1/2}h^{-\beta-1} \vee 1)\Big). \qquad (4.44)
\end{aligned}
$$

Hence, when $\alpha \geqslant 1/2$ the asymptotically optimal choice $h = (\frac{\log n}{n} \wedge \frac{1}{m})^{1/(2\alpha+2(\beta\vee 1/2)+1)}$ yields

$$
\Big| \int_{-\infty}^{q_\tau} (\widetilde{f}_h(x) - f(x))\mathrm{d}x \Big| = \mathcal{O}_{P,\mathcal{C}^\alpha\times\mathcal{D}^\beta}\Big((\tfrac{\log n}{n} \vee \tfrac{1}{m})^{(\alpha+1)/(2\alpha+2\beta+1)} \vee (\tfrac{\log n}{n} \vee \tfrac{1}{m})^{1/2}\Big).
$$

The result (4.44) can be obtained as Proposition 4.5 except for the term $T_{s,x} = T_{s,x}(q_\tau)$, defined in (4.30), which will be treated in the following. Defining the grid $\tau_1 = \sigma_0 \leqslant \ldots \leqslant \sigma_M = \tau_2$ such that $q_{\sigma_{k+1}} - q_{\sigma_k} \leqslant (q_{\tau_2} - q_{\tau_1})/M$ for $k = 1,\ldots,M$ and $M \in \mathbb{N}$, we decompose for any $c > 0$

$$
\begin{aligned}
P\Big(\sup_{\tau\in(\tau_1,\tau_2)} |T_{s,x}(q_\tau)| > c\Big) \leqslant& P\Big(\max_{k=1,\ldots,M} |T_{s,x}(q_{\sigma_k})| > c/2\Big) \qquad (4.45) \\
&+ P\Big(\sup_{\substack{q_1,q_2\in(q_{\tau_1},q_{\tau_2}):\\ |q_1-q_2|\leqslant(q_{\tau_2}-q_{\tau_1})/(2M)}} |T_{s,x}(q_1) - T_{s,x}(q_2)| > c/2\Big).
\end{aligned}
$$

For the first term we deduce a concentration inequality. We write

$$
\frac{1}{2\pi}T_{s,x} = \frac{1}{2\pi}T_{s,x}(q_\tau) = \frac{1}{n}\sum_{j=1}^n (\xi_{j,h}(q_\tau) - \mathbb{E}[\xi_{j,h}(q_\tau)])
$$

with

$$\xi_{j,h}(q_\tau) = \int_{-\infty}^{0} a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x+q_\tau)\mathrm{d}x = \mathcal{F}^{-1}\Big[\mathcal{F}\,a_s(-u)\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](q_\tau).$$

Uniformly in q_τ we have the deterministic bound

$$|\xi_{j,h}(q_\tau)| \leqslant \frac{1}{2\pi}\int_{-1/h}^{1/h} |\mathcal{F}\,a_s(-u)|\Big|\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}\Big|\mathrm{d}u \lesssim \int_{-1/h}^{1/h}\frac{1}{(1+|u|)|\varphi_\varepsilon(u)|}\mathrm{d}u \lesssim h^{-\beta} \quad (4.46)$$

Hence, $|\xi_{j,h}(q_\tau) - \mathbb{E}[\xi_{j,h}(q_\tau)]| \lesssim h^{-\beta}$. Since the variance of $T_{s,x}(q_\tau)$ is bounded by (4.31), Bernstein's inequality (e.g. Massart (2007), Prop. 2.9) yields for some constant $C > 0$ independent of q_τ

$$P(|T_{s,x}(q_\tau)| \geqslant \kappa(n^{-1/2}h^{-\beta+1/2} \vee n^{-1/2})) \leqslant 2\exp\Big(-\frac{C\kappa^2}{1+\kappa(nh)^{-1/2}}\Big).$$

For the second term on the right-hand side of (4.45) we estimate

$$\begin{aligned}
|T_{s,x}(q_1) - T_{s,x}(q_2)| &\leqslant \Big\|\Big(\mathcal{F}^{-1}\Big[\mathcal{F}\,a_s(u)\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}(\varphi_n(u)-\varphi_Y(u))\Big]\Big)'\Big\|_\infty |q_1 - q_2| \\
&\leqslant \frac{|q_1-q_2|}{2\pi}\int |u||\mathcal{F}\,a_s(u)|\frac{|\varphi_K(hu)|}{|\varphi_\varepsilon(u)|}|\varphi_n(u)-\varphi_Y(u)|\mathrm{d}u \\
&\lesssim |q_1-q_2|\int_{-1/h}^{1/h}(1+|u|)^\beta|\varphi_n(u)-\varphi_Y(u)|\mathrm{d}u.
\end{aligned}$$

Using Markov's inequality, we thus estimate (4.45) by

$$\begin{aligned}
&P\Big(\sup_{\tau\in(\tau_1,\tau_2)}|T_{s,x}(q_\tau)| > \kappa(n^{-1/2}h^{-\beta+1/2}\vee n^{-1/2})\Big) \\
&\quad\lesssim M\exp\Big(-\frac{C\kappa^2}{4+2\kappa(nh)^{-1/2}}\Big) \\
&\qquad + \frac{(q_{\tau_2}-q_{\tau_1})n^{1/2}(h^{\beta-1/2}\wedge 1)}{M\kappa}\,\mathbb{E}\Big[\int_{-1/h}^{1/h}(1+|u|)^\beta|\varphi_n(u)-\varphi_Y(u)|\mathrm{d}u\Big] \\
&\quad\lesssim M\exp\Big(-\frac{C\kappa^2}{4+2\kappa(nh)^{-1/2}}\Big) + \frac{(q_{\tau_2}-q_{\tau_1})(h^{-3/2}\wedge(h^{-\beta-1}))}{M\kappa}.
\end{aligned}$$

Choosing $M = n^2$ and $\kappa = (\frac{9}{C}\log n)^{1/2}$, we have $\kappa(nh)^{-1/2} = o(1)$ and the previous display converges to zero. Hence,

$$\sup_{\tau\in(\tau_1,\tau_2)}|T_{s,x}(q_\tau)| = \mathcal{O}_P\big((\tfrac{\log n}{n})^{1/2}h^{-\beta+1/2}\vee(\tfrac{\log n}{n})^{1/2}\big). \qquad \square$$

Proof of Theorem 4.10

To prove the lower bound for the estimation of the distribution function we can assume without loss of generality $q = 0$. For $n \leqslant m$ the estimation error of $\bar{F}_{n,m}(0)$ is bounded from below by the estimation error with known error distribution. A lower bound for the latter is proved by Fan (1991b) whose construction can be used in our setting, too.

To prove the lower bound for $m < n$, we will apply Theorem 2.1 in Tsybakov (2009). To this end, we construct two alternatives $(F_i, f_{\varepsilon,i}) \in \widetilde{\mathcal{C}}^\alpha(R, r, [-\zeta, \zeta]) \times \mathcal{D}^\beta(R, \gamma)$, $i = 1, 2$, such that the χ^2-distance of the corresponding laws of $(Y_1, \dots, Y_n, \varepsilon_1^*, \dots \varepsilon_m^*)$ is bounded by some small constant and such that $|F_1(0) - F_2(0)|$ is bounded from below with the right rate. Recall that the convolution of a c.d.f. F with a function g is defined as $F * g(x) = \int g(x-y)\mathrm{d}F(y)$. Following the idea by Neumann (1997) our construction will satisfy $F_1 * f_{\varepsilon,1} = F_2 * f_{\varepsilon,2}$ and is thus independent of n.

Step 1: For the construction of the alternatives we need the following: Let f_0 be a bounded density whose corresponding distribution is in $\mathcal{C}^\alpha(R, r, \zeta)$ satisfying $q_\tau = 0$. Let $f_{\varepsilon,0}$ be an inner point of $\mathcal{D}^\beta(R, \gamma)$ with

$$f_{\varepsilon,0}(x) \gtrsim (1 + |x|)^{-\gamma-2}, \quad |(\mathcal{F} f_{\varepsilon,0})^{(k)}(u)| \lesssim (1 + |u|)^{-\beta}, k = 0, \dots, K \tag{4.47}$$

for $x, u \in \mathbb{R}$ and an integer $K > \gamma/2 + 1$. Let the perturbation $g \in C^\infty(\mathbb{R}) \cap L^1(\mathbb{R})$ satisfy

$$\int g(x)\mathrm{d}x = 0, \int_{-\infty}^0 g(x)\mathrm{d}x \neq 0, \|(1 \vee x^{\gamma \vee 1})g(x)\|_{L^1} < \infty, \operatorname{supp} \mathcal{F} g \subseteq [-2, -1] \cup [1, 2]. \tag{4.48}$$

Define $g_h := h^{-1}g(\bullet/h)$ for $h > 0$ and for some $a \in (0, 1), c > 0$

$$\begin{aligned} F_1(x) &:= a \int_{-\infty}^x f_0(y)\mathrm{d}y + (1 - a)\mathbb{1}_{[2\zeta,\infty)}(x), \\ f_{\varepsilon,1}(x) &:= f_{\varepsilon,0} + ch^{\alpha+1}(f_{\varepsilon,0} * g_h(\bullet + 2\zeta))(x), \\ F_2(x) &:= F_1(x) + ch^{\alpha+1} \int_{-\infty}^x g_h(\bullet + 2\zeta) * F_1(y)\mathrm{d}y, \\ f_{\varepsilon,2}(x) &:= f_{\varepsilon,0}(x). \end{aligned} \tag{4.49}$$

Owing to $\int g_h = 0$, F_i are distribution functions admitting Lebesgue densities on $[-\zeta, \zeta]$ which are at least α-Hölder continuous. Estimating $\|f_0 * g_h\|_{C^\alpha(\mathbb{R})} \lesssim \|f_0\|_{L^1}\|g_h\|_{C^\alpha(\mathbb{R})} \lesssim h^{-\alpha-1}$, we infer that $\mathrm{d}F_2$ is contained in a closed Hölder ball. Hence, we obtain $F_i \in \widetilde{\mathcal{C}}^\alpha(R, r, [-\zeta, \zeta])$ for $c > 0$ sufficiently small. $f_{\varepsilon_i} \in \mathcal{D}^\beta(R, \gamma)$ can be verified, using $\int g = 0, \|\mathcal{F} g\|_\infty \leqslant \|g\|_{L^1}$ and $\|(\mathcal{F} g)'(u)(1 + |u|)\|_\infty < \infty$.

Step 2: To bound the distance $|F_1(0) - F_2(0)|$ from below we note, using Fubini's theorem, $\int g = 0$ and $\|f_0\|_\infty < \infty$,

$$\begin{aligned} F_2(0) - F_1(0) &= h^{\alpha+1}\Big(ac \int \int_{2\zeta}^{-y+2\zeta} f_0(x)g_h(y)\mathrm{d}x\mathrm{d}y + (1 - a)c \int_{-\infty}^0 g_h(x)\mathrm{d}x\Big) \\ &= h^{\alpha+1}\Big((1 - a)c \int_{-\infty}^0 g(x)\mathrm{d}x + \mathcal{O}(\|yg_h(y)\|_{L^1})\Big) \\ &= h^{\alpha+1}\Big((1 - a)c \int_{-\infty}^0 g(x)\mathrm{d}x + \mathcal{O}(h)\Big), \end{aligned} \tag{4.50}$$

for h small enough. Therefore, $|F_1(0) - F_2(0)| \gtrsim h^{\alpha+1}$.

Step 3: Using the independence of the observations, the sample $(Y_1, \dots, Y_n, \varepsilon_1^*, \dots \varepsilon_m^*)$ is distributed according to $(F_i * f_{\varepsilon,i})^{\otimes n} \otimes f_{\varepsilon,i}^{\otimes m}$ under the hypotheses $i = 1, 2$. By construction $F_1 * f_{\varepsilon,1} = F_2 * f_{\varepsilon,2}$ such that the χ^2-distance of the laws of the observations equals

$$\chi^2(f_{\varepsilon,1}^{\otimes m}, f_{\varepsilon,2}^{\otimes m}) = \Big(1 + \int \frac{(f_{\varepsilon,1} - f_{\varepsilon,2})^2(x)}{f_{\varepsilon,1}(x)}\mathrm{d}x\Big)^m - 1. \tag{4.51}$$

We decompose

$$\begin{aligned}
&\int \frac{(f_{\varepsilon,1} - f_{\varepsilon,2})^2(x)}{f_{\varepsilon,2}(x)}\mathrm{d}x \\
=&c^2h^{2\alpha+2}\Big(\int_{|x|\leqslant 1} \frac{(f_{\varepsilon,0} * g_h(\bullet + 2\zeta))^2(x)}{f_{\varepsilon,0}(x)}\mathrm{d}x + \int_{|x|>1} \frac{(f_{\varepsilon,0} * g_h(\bullet + 2\zeta))^2(x)}{f_{\varepsilon,0}(x)}\mathrm{d}x\Big) \\
=:&c^2h^{2\alpha+2}(I_1 + I_2).
\end{aligned}$$

For the first integral we use $\inf_{|x|\leqslant 1} f_{\varepsilon,0}(x) > 0$, Plancherel's identity, $f_{\varepsilon,0} \in \mathcal{D}^\beta(R,\gamma)$ and the support of $\mathcal{F}\, g$ to estimate

$$|I_1| \lesssim \int |\mathcal{F}\, f_{\varepsilon,0}(u)\,\mathcal{F}\, g(hu)e^{-i2\zeta u}|^2\mathrm{d}u \lesssim \int_{1/h\leqslant|u|\leqslant 2/h} (1+|u|)^{-2\beta}\mathrm{d}u \lesssim h^{2\beta-1}.$$

Using (4.47), I_2 can be estimated similarly

$$\begin{aligned}
|I_2| &\lesssim \int_{|x|>1} (1+|x|)^{\gamma+2}|x|^{-2K}|\,\mathcal{F}^{-1}[(\mathcal{F}\, f_{\varepsilon,0}\,\mathcal{F}\, g_h e^{-i2\zeta\bullet})^{(K)}]|^2(x)\mathrm{d}x \\
&\sim \int_{1/h\leqslant|u|\leqslant 2/h} |(\mathcal{F}\, f_{\varepsilon,0}(u)\,\mathcal{F}\, g(hu)e^{-i2\zeta u})^{(K)}|^2\mathrm{d}u \lesssim h^{2\beta-1}.
\end{aligned}$$

We conclude from (4.51) for some constant $C > 0$ that

$$\chi^2(f_{\varepsilon,1}^{\otimes m}, f_{\varepsilon,2}^{\otimes m}) \leqslant (1 + Cc^2h^{2\alpha+2\beta+1})^m - 1 \leqslant \exp(Cc^2mh^{2\alpha+2\beta+1}) - 1,$$

which can be bounded by an arbitrarily small constant if c is chosen sufficiently small and $h = m^{-1/(2\alpha+2\beta+1)}$. We obtain from Step 2 that there is some positive constant C such that $|F_1(0) - F_2(0)| \geq Cm^{-(\alpha+1)/(2\alpha+2\beta+1)}$.

Step 4: Replacing in (4.49) the factor $h^{\alpha+1}$ in F_2 and $f_{\varepsilon,1}$ by $cm^{-1/2}$ for some sufficiently small constant $c > 0$ and choosing $h = 1$, the previous steps yield the lower bound $m^{-1/2}$.

Let us finally conclude the lower bound for the estimation error of the quantiles. We use the construction from Step 1, denoting the τ-quantile of F_i by $q_{\tau,i}$. We note $|q_{\tau,1}| < \delta$ for any $\delta > 0$ if we choose a close enough to one and thus F_1 is regular in an interval around $q_{\tau,1}$. Moreover, it holds

$$\begin{aligned}
\|F_1 - F_2\|_\infty &\leqslant c(m^{-1/2} \vee h^{\alpha+1})\big\|(af_0 + (1-a)\delta_{-2\zeta}) * g_h(\bullet + 2\zeta)\big\|_{L^1} \\
&\leqslant c(m^{-1/2} \vee h^{\alpha+1})\|g\|_{L^1} \to 0.
\end{aligned}$$

We infer analogously to (4.40) that $|q_{\tau,1} - q_{\tau,2}| < \delta$ for any $\delta > 0$ and m sufficiently large implying $F_i \in \widetilde{\mathcal{C}}^\alpha(R,r,\zeta)$. Applying a Taylor expansion similar to (4.9), we obtain

$$q_{\tau,2} - q_{\tau,1} = -\frac{F_2(q_{\tau,1}) - F_1(q_{\tau,1})}{F_2'(q_\tau^*)}$$

for some intermediate point between $q_{\tau,1}$ and $q_{\tau,2}$. The denominator $F_2'(q_\tau^*)$ is bounded from above and below owing to $\sup_{|x|\leqslant\zeta}|F_2'(x)-af_0(x)|\to 0$, $|q_{\tau,2}|\leqslant|q_{\tau,2}-q_{\tau,1}|+|q_{\tau,1}|<2\delta$ and $f_0(0)>0$. (4.50) yields $|q_{\tau,2}-q_{\tau,1}|\gtrsim m^{-1/2}\vee h^{\alpha+1}$. The assertion follows from Steps 3 and 4 above. □

4.4.2. Proofs for Section 4.2

We start with Lemma 4.11 concerning the bandwidth set $\mathcal{B}_n$ from (4.14).

Proof of Lemma 4.11

Property (i) is satisfied by construction. For (ii) we note that

$$N_n\sim(\log n-3\log\log n)/\log L\lesssim\log n$$

and $(\log n)^2 h_{N_n}\sim(\log n)^{-1}$. It remains to verify $nh_{\widetilde{j}_n}^{2\beta+2}\to\infty$ and property (iii).

By Lemma 4.13 we can argue on the event $B_\varepsilon(h)$ from (4.6). The deterministic counterpart of $\widetilde{j}_n$, defined in (4.13), is given by

$$j_{0,n}:=\min\Big\{j=0,\dots,N_n:2\leqslant\Big(\frac{\log n}{n}\Big)^{1/2}\int_{-1/h_j}^{1/h_j}|\varphi_\varepsilon(u)|^{-1}\mathrm{d}u\leqslant 4\Big\}. \tag{4.52}$$

Noting that for $f_\varepsilon\in\mathcal{D}^\beta(R,\gamma)$

$$4\geqslant\Big(\frac{\log n}{n}\Big)^{1/2}\int_{-1/h_{j_{0,n}}}^{1/h_{j_{0,n}}}|\varphi_\varepsilon(u)|^{-1}\mathrm{d}u\gtrsim\Big(\frac{\log n}{nh_{j_{0,n}}^{2\beta+2}}\Big)^{1/2}$$

we obtain $nh_{j_{0,n}}^{2\beta+2}\to\infty$ and thus it is sufficient to prove

$$\inf_{f\in\mathcal{C}^\alpha(R,r,\zeta)}\inf_{f_\varepsilon\in\mathcal{D}^\beta(R,\gamma)}P\big(\{h_{j_{0,n}}\leqslant h_{\widetilde{j}_n}\leqslant h^*\}\cap B_\varepsilon(h_{j_{0,n}})\big)\to 1\quad\text{as}\quad n\to\infty. \tag{4.53}$$

for the optimal bandwidth $h^*=n^{-1/(2\alpha+2(\beta\vee 1/2)+1)}$. For convenience we define

$$I_n(h):=\Big(\frac{\log n}{n}\Big)^{1/2}\int_{-1/h}^{1/h}\frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|},\quad\widetilde{I}_n(h):=\Big(\frac{\log n}{n}\Big)^{1/2}\int_{-1/h}^{1/h}\frac{\mathrm{d}u}{|\varphi_{\varepsilon,m}(u)|}.$$

Assume $h_{\widetilde{j}_n}<h_{j_{0,n}}$, then monotonicity implies $\widetilde{I}_n(h_{j_{0,n}})\leqslant\widetilde{I}_n(h_{\widetilde{j}_n})\leqslant 1$. Combined with $I_n(j_{0,n})\geqslant 2$, we obtain $I_n(h_{j_{0,n}})-\widetilde{I}_n(h_{j_{0,n}})\geqslant 1$. Hence,

$$\{h_{\widetilde{j}_n}<h_{j_{0,n}}\}\subseteq\{|I_n(h_{j_{0,n}})-\widetilde{I}_n(h_{j_{0,n}})|\geqslant 1\}. \tag{4.54}$$

On the other hand, if $h^*<h_{\widetilde{j}_n}$, we get $\widetilde{I}_n(h^*)\geqslant\widetilde{I}_n(h_{\widetilde{j}_n})\geqslant 1/2$. Since $I_n(h^*)\lesssim(\frac{\log n}{n(h^*)^{2\beta+2}})^{1/2}$ converges to zero, $I_n(h^*)\leqslant 1/4$ for n large enough. Thus,

$$\{h_{\widetilde{j}_n}>h^*\}\subseteq\{|I_n(h^*)-\widetilde{I}_n(h^*)|\geqslant 1/4\}. \tag{4.55}$$

To show that the probabilities of the right-hand sides of (4.54) and (4.55) converge to zero, we first apply the Cauchy–Schwarz inequality

$$\begin{aligned}\big|I_n(h)-\widetilde{I}_n(h)\big|^2 &\leqslant \frac{\log n}{n}\int_{-1/h}^{1/h}\frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2}\int_{-1/h}^{1/h}\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathrm{d}u\\ &\lesssim \frac{\log n}{nh^{2\beta+1}}\int_{-1/h}^{1/h}\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathrm{d}u.\end{aligned}$$

Markov's inequality and (4.33) yield for $h\in\{h_{min},h^*\}$

$$\begin{aligned}P\Big(\{\big|I_n(h)-\widetilde{I}_n(h)\big|\geqslant\tfrac14\}\cap B_\varepsilon(h_{j_{0,n}})\Big) &\lesssim \frac{\log n}{nh^{2\beta+1}}\int_{-1/h}^{1/h}\mathbb{E}\Big[\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathbb{1}_{B_\varepsilon(h_{j_{0,n}})}\Big]\mathrm{d}u\\ &\lesssim \frac{\log n}{nmh^{4\beta+2}}\end{aligned}$$

which converges to zero. Therefore, (4.53) holds true. □

Before we can prove Theorem 4.12, some preparations are needed. By Lemma 4.14 there is a constant $D>0$ such that the bias can be bounded by $B_h:=Dh^{\alpha+1}$. By the error representation (4.42) we have for any $h\in\mathcal{B}$

$$\begin{aligned}|\widetilde{q}_{\tau,h}-q_\tau| &= \Big|\frac{\int_{-\infty}^{q_\tau}(\widetilde{f}_h(x)-f(x))\mathrm{d}x-\widetilde{M}_h(\widetilde{q}_{\tau,h})}{\widetilde{f}_h(\widetilde{q}^*)}\Big|\\ &\leqslant \frac{B_h+|V_{h,X}+V_{h,\varepsilon}+V_{h,c}|+|\widetilde{M}_h(\widetilde{q}_{\tau,h})|}{|\widetilde{f}_h(q^*)|}\end{aligned}\tag{4.56}$$

with some $q^*\in[(q_\tau\wedge\widetilde{q}_{\tau,h}),(q_\tau\vee\widetilde{q}_{\tau,h})]$ and where the stochastic error is decomposed in

$$V_{h,X}:=\frac1n\sum_{j=1}^n\left(\xi_j(h)-\mathbb{E}[\xi_j(h)]\right)\quad\text{with}\tag{4.57}$$

$$\xi_j(h):=\int_{-\infty}^0 a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x+q_\tau)\mathrm{d}x,$$

$$V_{h,\varepsilon}:=\int_{-\infty}^0 a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)\varphi_n(u)}{\varphi_\varepsilon(u)}\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big)\Big](x+q_\tau)\mathrm{d}x\tag{4.58}$$

$$V_{h,c}:=\int_{-\infty}^0 a_c(x)\,\mathcal{F}^{-1}\Big[\varphi_K(hu)\Big(\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)}-\varphi_X(u)\Big)\Big](x+q_\tau)\mathrm{d}x.\tag{4.59}$$

In view of the analysis in Section 4.4.1, the part of the stochastic error which is due to the continuous part a_c will be negligible. Hence, we concentrate on $V_{h,X}$ and $V_{h,\varepsilon}$. By independence of $(\xi_j(h))_j$, we obtain

$$\operatorname{Var}(V_{h,X})\leqslant\frac1n\mathbb{E}[\xi_j(h)^2]=\frac1n\mathbb{E}\Big[\Big(\int_{-\infty}^0 a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x+q_\tau)\mathrm{d}x\Big)^2\Big]=:\sigma_{h,X}^2.\tag{4.60}$$

We will determine the variance of $V_{h,\varepsilon}$ on the event $B_\varepsilon(h)$, defined in (4.6). We apply Plancherel's identity and the Cauchy–Schwarz inequality to separate Y_i and ε_i from each other:

$$\begin{aligned}
\mathbb{E}[|V_{h,\varepsilon}|\mathbb{1}_{B_\varepsilon(h)}] =& \frac{1}{2\pi}\,\mathbb{E}\Big[\Big|\int \mathcal{F}\,a_s(-u)e^{-iuq_\tau}\frac{\varphi_K(hu)\varphi_n(u)}{\varphi_\varepsilon(u)}\Big(\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big)\mathrm{d}u\Big|\mathbb{1}_{B_\varepsilon(h)}\Big] \\
\leqslant& \frac{1}{2\pi}\,\mathbb{E}\Big[\Big(\int |\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\Big|^2\mathrm{d}u\Big)^{1/2} \\
&\times\Big(\int |\varphi_K(hu)||\mathcal{F}\,a_s(-u)|^2\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|^2\mathrm{d}u\Big)^{1/2}\mathbb{1}_{B_\varepsilon(h)}\Big] \\
\leqslant& \frac{1}{2\pi}\,\mathbb{E}\Big[\Big(\int |\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\Big|^2\mathrm{d}u\Big)^{1/2}\Big(\int |\varphi_K(hu)|\Big|\frac{\mathcal{F}\,a_s(-u)}{\varphi_{\varepsilon,m}(u)}\Big|^2\mathrm{d}u\Big)^{1/2} \\
&\times\sup_{|u|\leqslant 1/h}|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|\mathbb{1}_{B_\varepsilon(h)}\Big]. \qquad (4.61)
\end{aligned}$$

Let us define

$$\sigma_{h,\varepsilon} := \frac{1}{2\pi}m^{-1/2}\sigma_{h,\varepsilon,1}\sigma_{h,\varepsilon,2} \qquad (4.62)$$

with

$$\begin{aligned}
\sigma_{h,\varepsilon,1} &:= \mathbb{E}\Big[\Big(\int |\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\Big|^2\mathrm{d}u\Big)^{1/2}\Big], \\
\sigma_{h,\varepsilon,2} &:= \mathbb{E}\Big[\Big(\int |\varphi_K(hu)|\Big|\frac{\mathcal{F}\,a_s(-u)}{\varphi_{\varepsilon,m}(u)}\Big|^2\mathrm{d}u\Big)^{1/2}\mathbb{1}_{B_\varepsilon(h)}\Big].
\end{aligned}$$

With the bounds $\sigma_{h,X}$ and $\sigma_{h,\varepsilon}$ at hand, we obtain the following concentration results.

Lemma 4.17. *Let $\mathcal{B}$ be a set satisfying $|\mathcal{B}| \lesssim \log n, (\log\log n)/nh_1 \to 0$ for $h_1 = \min\mathcal{B}$ as well as $|\log h_1| \lesssim \log n$. Then we obtain uniformly over $f \in \mathcal{C}^\alpha(R,r,\zeta)$ and $f_\varepsilon \in \mathcal{D}^\beta(R,\gamma)$ for any $\delta > 0$:*

(i) $P\Big(\exists h\in\mathcal{B} : |V_{h,X}| \geqslant (1+\delta)\sqrt{\log\log n}\big(\sqrt{2}\sigma_{h,X} + o(n^{-1/2}(h^{-\beta+1/2}\vee 1))\big)\Big) \to 0.$

(ii) $P\Big(\exists h\in\mathcal{B} : |V_{h,\varepsilon}| \geqslant \delta(\log n)^3\sigma_{h,\varepsilon}\Big) \to 0.$

(iii) Assuming further $mh_1^{(2\beta\wedge 1)+2} \gtrsim 1$,
$P\Big(\exists h\in\mathcal{B} : |V_{h,c}| \geqslant (\log n)^{3/2}n^{-1/2}(h^{-\beta+1/2}\vee 1)\Big) \to 0.$

Proof. *(i)* Using the deterministic bound (4.46), we obtain $|\xi_j(h) - \mathbb{E}[\xi_j(h)]| \leqslant Ch^{-\beta}$ for some constant $C > 0$. Since the variance is bounded by (4.60), Bernstein's inequality (e.g. Massart (2007), Prop. 2.9) yields for any positive $\kappa_n = o(nh)$

$$P\Big(|V_{h,X}| \geqslant \sqrt{2\sigma_{h,X}^2\kappa_n} + \frac{C\kappa_n}{3nh^\beta}\Big) \leqslant 2e^{-\kappa_n}.$$

Hence, $\sqrt{\kappa_n}(nh^\beta)^{-1} \lesssim (n(h^{2\beta-1}\wedge 1))^{-1/2}\big(\kappa_n/(nh)\big)^{1/2}$ yields uniformly in $\mathcal{C}^\alpha(R,r,\zeta)$ and $\mathcal{D}^\beta(R,\gamma)$

$$P(|V_{h,X}| \geqslant \sqrt{\kappa_n}(\sqrt{2}\sigma_{h,X} + o(n^{-1/2}(h^{-\beta+1/2}\vee 1)))) \leqslant 2e^{-\kappa_n}$$

(ii) Using an estimate as in (4.61), we obtain

$$\begin{aligned}
|V_{h,\varepsilon}| &\leqslant \frac{1}{2\pi}\Big(\int |\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\Big|^2 \mathrm{d}u\Big)^{1/2} \\
&\qquad\times \Big(\int |\varphi_K(hu)||\mathcal{F}\, a_s(-u)|^2\Big|\frac{\varphi_\varepsilon(u)-\varphi_{\varepsilon,m}(u)}{\varphi_{\varepsilon,m}(u)}\Big|^2 \mathrm{d}u\Big)^{1/2} \\
&\leqslant \frac{1}{2\pi}\Big(\int |\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\Big|^2 \mathrm{d}u\Big)^{1/2}\Big(\int |\varphi_K(hu)|\Big|\frac{\mathcal{F}\, a_s(-u)}{\varphi_{\varepsilon,m}(u)}\Big|^2 \mathrm{d}u\Big)^{1/2} \\
&\qquad\times \sup_{|u|\leqslant 1/h} |\varphi_\varepsilon(u)-\varphi_{\varepsilon,m}(u)| \\
&=: \frac{1}{2\pi} V_{h,\varepsilon,1}V_{h,\varepsilon,2} \sup_{|u|\leqslant 1/h}|\varphi_\varepsilon(u)-\varphi_{\varepsilon,m}(u)|.
\end{aligned}$$

Hence for any $c \in (0, 1/4)$

$$\begin{aligned}
&P\Big(\{|V_{h,\varepsilon}| \geqslant \delta(\log n)^3\sigma_{h,\varepsilon}\} \cap B_\varepsilon(h)\Big) \\
&\quad\leqslant P(|V_{h,\varepsilon,1}| \geqslant (\log n)^{1+c}\sigma_{h,\varepsilon,1}) + P\Big(\{|V_{h,\varepsilon,2}| \geqslant (\log n)^{1+c}\sigma_{h,\varepsilon,2}\} \cap B_\varepsilon(h)\Big) \\
&\qquad + P\Big(\sup_{|u|\leqslant 1/h}|\varphi_\varepsilon(u)-\varphi_{\varepsilon,m}(u)| \geqslant \delta(\log n)^{1-2c}m^{-1/2}\Big) =: P_{h,1}+P_{h,2}+P_{h,3}.
\end{aligned}$$

The first two probabilities can be bounded by Markov's inequality:

$$\begin{aligned}
P_{h,1} &\leqslant (\log n)^{-1-c}\sigma_{h,\varepsilon,1}^{-1}\,\mathbb{E}[V_{h,\varepsilon,1}] = (\log n)^{-1-c}, \\
P_{h,2} &\leqslant (\log n)^{-1-c}\sigma_{h,\varepsilon,2}^{-1}\,\mathbb{E}[V_{h,\varepsilon,2}\mathbb{1}_{B_\varepsilon(h)}] = (\log n)^{-1-c}.
\end{aligned}$$

For $P_{h,3}$ we will apply the following version of Talagrand's inequality (cf. Massart (2007), (5.50)): Let T be a countable index and for all $t \in T$ let $Z_{1,t}, \dots, Z_{n,t}$ be an i.i.d. sample of centered, complex valued random variables satisfying $\|Z_{k,t}\|_\infty \leqslant h$, for all $t \in T, k = 1, \dots, n$, as well as $\sup_{t\in T}\operatorname{Var}(\sum_{k=1}^n Z_{k,t}) \leqslant v < \infty$. Then for all $\kappa > 0$

$$P\Big(\sup_{t\in T}\Big|\sum_{k=1}^n Z_{k,t}\Big| \geqslant 4\,\mathbb{E}\Big[\sup_{t\in T}\Big|\sum_{k=1}^n Z_{k,t}\Big|\Big] + \sqrt{2v\kappa} + \frac{2}{3}h\kappa\Big) \leqslant 2e^{-\kappa}. \tag{4.63}$$

Choosing the rational numbers $T = \mathbb{Q}\cap[-\frac{1}{h}, \frac{1}{h}]$ and $Z_{k,t} := e^{it\varepsilon_k^*} - \varphi_\varepsilon(t)$, Talagrand's inequality applies with $h = 2$ and $v = n$. As in (4.21) we use Theorem 4.1 by Neumann and Reiß (2009) to obtain for any $\eta \in (0, 1/2)$

$$m^{1/2}\,\mathbb{E}\Big[\sup_{|u|\leqslant 1/h}|\varphi_{\varepsilon,m}(t)-\varphi_\varepsilon(t)|\Big] \lesssim |\log h|^{1/2+\eta}.$$

Therefore on the assumptions $\kappa_n^{-1}(\log n)^{1+2\eta} \to 0$ and $\kappa_n/m \to 0$

$$4\,\mathbb{E}\Big[\sup_{|u|\leqslant 1/h, u\in\mathbb{Q}}|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|\Big] + \sqrt{\frac{2\kappa_n}{m}} + \frac{4}{3m}\kappa_n = \sqrt{\frac{\kappa_n}{m}}(\sqrt{2}+o(1))$$

and thus continuity of $\varphi_{\varepsilon,m}$ and (4.63) yield

$$P_{h,3} = P\Big(\sup_{|u|\leqslant 1/h, u\in\mathbb{Q}} |\varphi_{\varepsilon,m}(u) - \varphi_\varepsilon(u)| \geqslant (\sqrt{2}+o(1))\sqrt{\kappa_n/m}\Big) \leqslant 2e^{-\kappa_n}. \quad (4.64)$$

With $\kappa_n = \frac{\delta}{2}(\log n)^{2-4c}$ for $c < 1/4 - \eta/2$ we obtain $P_3 \leqslant 2n^{-\delta/2}$. Using $h_1 = \min \mathcal{B}$ and $|\log B| \lesssim \log n$ as well as Lemma 4.13, we finally get

$$P\Big(\sup_{h\in\mathcal{B}} |V_{h,\varepsilon}| \geqslant (\sqrt{2}+\delta)(\log n)^3 \sigma_{h,\varepsilon}\Big) \leqslant \sum_{h\in\mathcal{B}} (P_{h,1} + P_{h,2} + P_{h,3}) + P(B_\varepsilon(h_1)^c) = o(1).$$

(iii) Corollary 4.15 shows for $\delta_h > 0$ and for any sequence $(x_n)_n$ that tends to infinity

$$P\Big(\exists h \in \mathcal{B} : |V_{h,c}| \geqslant \delta_h\Big) \lesssim \sum_{h\in\mathcal{B}} \frac{x_n}{\delta_h^2 n(mh^{2\beta+2} \wedge 1)} + o(1).$$

Choosing $\delta_h = (\log n)^{3/2} n^{-1/2}(h^{-\beta+1/2} \vee 1)$ and $x_n = o((\log n)^{1/2})$ yields

$$\begin{aligned} &P\Big(\exists h \in \mathcal{B} : |V_{h,c}| \geqslant (\log n)^{3/2} n^{-1/2}(h^{-\beta+1/2} \vee 1)\Big) \\ &\quad\lesssim \sum_{h\in\mathcal{B}} \frac{x_n}{(\log n)^3 (mh^{(2\beta\wedge 1)+2} \wedge 1)} + o(1) \lesssim \frac{x_n}{(\log n)^2 (mh^{(2\beta\wedge 1)+2} \wedge 1)} + o(1) = o(1). \end{aligned}$$

□

For the denominator in the error representation (4.56) we need uniform consistency. A uniform result on the error $|\widetilde{q}_{\tau,h} - q_\tau|$ follows immediately.

Lemma 4.18. *Let $\mathcal{B}$ be a finite set satisfying $|\mathcal{B}| \lesssim \log n$, $\sup_{h\in\mathcal{B}} h\log(n) \to 0$ as well as $\sup_{h\in\mathcal{B}} (\log n)^2/(nh^{2\beta+1}) \to 0$. Then we obtain for $n \to \infty$ and $\eta \in (0,1)$*

$$\sup_{f\in\mathcal{C}^\alpha(R,r,\zeta,U_n)} \sup_{f_\varepsilon\in\mathcal{D}^\beta(R,\gamma)} P\Big(\sup_{h\in\mathcal{B}} \sup_{q_\tau^*\in[q_\tau\wedge\widetilde{q}_{\tau,h}, q_\tau\vee\widetilde{q}_{\tau,h}]} |\widetilde{f}_h(q_\tau^*) - f(q_\tau)| > \eta f(q_\tau)\Big) \to 0. \quad (4.65)$$

Moreover, supposing $\min_{h\in\mathcal{B}} nh^{(2\beta\wedge 1)+2} \gtrsim 1$, we obtain uniformly in $f \in \mathcal{C}^\alpha(R,r,\zeta)$ and $f_\varepsilon \in \mathcal{D}^\beta(R,\gamma)$ for any sequence of critical values $(\delta_h)_{h\in\mathcal{B}}$ satisfying $\inf_{\mathcal{B}} \delta_h \to \infty$

$$P\Big(\exists h \in \mathcal{B} : |\widetilde{q}_{\tau,h} - q_\tau| > \delta_h\big(3Dh^{\alpha+1} + n^{-1/2}(h^{-\beta+1/2} \vee 1)\big)\Big) \lesssim \sum_{h\in\mathcal{B}} \frac{1}{\delta_h} + o(1). \quad (4.66)$$

Proof. Since $f(q_\tau) \geqslant r$ and $f \in C^\alpha([q_\tau - \zeta, q_\tau + \zeta], R)$, decomposition (4.43) implies with $\kappa = (\frac{\eta r}{2R})^{1\vee\alpha^{-1}} \wedge \zeta$

$$\begin{aligned} &P\Big(\sup_{h\in\mathcal{B}} \sup_{q_\tau^*\in[q_\tau\wedge\widetilde{q}_{\tau,h}, q_\tau\vee\widetilde{q}_{\tau,h}]} |\widetilde{f}_h(q_\tau^*) - f(q_\tau)| > \eta f(q_\tau)\Big) \\ \leqslant &P\big(\sup_{h\in\mathcal{B}} \sup_{x\in[-\kappa,\kappa]} |\widetilde{f}_h(x+q_\tau) - f(x+q_\tau)| > \eta r/2\big) + P(\sup_{h\in\mathcal{B}} |\widetilde{q}_{\tau,h} - q_h| > \kappa). \quad (4.67) \end{aligned}$$

Using $h_1 = \min \mathcal{B}$, the first probability can be bounded by

$$\begin{aligned}&\sum_{h\in\mathcal{B}} P\Big(\Big\{\sup_{x\in[-\kappa,\kappa]} |\widetilde{f}_h(x+q_\tau)-f(x+q_\tau)|>\eta r/2\Big\}\cap B_\varepsilon(h_1)\Big)+P(B_\varepsilon(h_1)^c)\\ \lesssim&\log n\sup_{h\in\mathcal{B}} P\Big(\Big\{\sup_{x\in[-\kappa,\kappa]} |\widetilde{f}_h(x+q_\tau)-f(x+q_\tau)|>\eta r/2\Big\}\cap B_\varepsilon(h_1)\Big)+o(1)=o(1),\end{aligned}$$

since for all h the probability in the last line converges faster to zero than $1/\log n$ owing to the concentration inequalities (4.35) and (4.36) and the conditions on h. To estimate the second term in (4.67), we apply Lemma 4.16. Therefore, the conditions $h\log(n)\to 0$ and $(\log n)^2/(nh^{2\beta+1})\to 0$ yield the first assertion.

The estimate (4.66) follows from the error decomposition (4.9), (4.65) and Corollary 4.15 with $x_n = o(\inf_{\mathcal{B}}\delta_h)$

$$\begin{aligned}&P\Big(\exists h\in\mathcal{B}:|\widetilde{q}_{\tau,h}-q_\tau|>\delta_h\big(3Dh^{\alpha+1}+n^{-1/2}(h^{-\beta+1/2}\vee 1)\big)\Big)\\ \leqslant&P\Big(\exists h\in\mathcal{B}:\Big|\int_{-\infty}^{q_\tau}\widetilde{f}_h(x)-f(x)\mathrm{d}x\Big|>\tfrac{1}{2}f(q_\tau)\delta_h\big(3Dh^{\alpha+1}+n^{-1/2}(h^{-\beta+1/2}\vee 1)\big)\Big)\\ &+P\Big(\sup_{h\in\mathcal{B}}\sup_{q_\tau^*\in[q_\tau\wedge\widetilde{q}_{\tau,h},q_\tau\vee\widetilde{q}_{\tau,h}]}|\widetilde{f}_h(q_\tau^*)-f(q_\tau)|>\tfrac{1}{2}f(q_\tau)\Big)\\ \lesssim&\sum_{h\in\mathcal{B}}\Big(\frac{1}{\delta_h}+\frac{1}{\delta_h^2}\frac{x_n}{mh^{1\wedge 2\beta+2}\wedge 1}\Big)+o(1)\lesssim\sum_{h\in\mathcal{B}}\frac{1}{\delta_h}+o(1).\end{aligned}$$ □

The variances $\sigma_{h,X}$ and $\sigma_{h,\varepsilon}$, defined in (4.60) and (4.62) can be estimated by

$$\begin{aligned}\widetilde{\sigma}_{h,X}^2=&\frac{1}{n^2}\sum_{j=1}^n\Big(\int_{-\infty}^0 a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_{\varepsilon,m}(u)}\Big](x+\widetilde{q}_{\tau,h})\mathrm{d}x\Big)^2,\\ \widetilde{\sigma}_{h,\varepsilon}^2=&\frac{1}{4\pi^2}m^{-1}\widetilde{\sigma}_{h,\varepsilon,1}^2\widetilde{\sigma}_{h,\varepsilon,2}^2\end{aligned}$$

with

$$\widetilde{\sigma}_{h,\varepsilon,1}^2=\int_{-1/h}^{1/h}|\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_{\varepsilon,m}(u)}\Big|^2\mathrm{d}u,\quad \widetilde{\sigma}_{h,\varepsilon,2}^2=\int_{-1/h}^{1/h}|\varphi_K(hu)|\frac{|\,\mathcal{F}a_s(u)|^2}{|\varphi_{\varepsilon,m}|^2}\mathrm{d}u.$$

The following two lemmas show that these estimators are indeed reasonable.

Lemma 4.19. *Let $\mathcal{B}$ be a finite set satisfying $|\mathcal{B}|\lesssim\log n$, $\max_{h\in\mathcal{B}}h^\alpha\log n\to 0$ as well as $\min_{h\in\mathcal{B}}nh^{2\beta+2}\to\infty$. Let $\widetilde{\sigma}_{h,X}$ and $\sigma_{h,X}$ be given in (4.16) and* (4.60) *respectively. Then we obtain for all $\eta>0$ as $n\to\infty$*

$$\sup_{f\in\mathcal{C}^\alpha(R,r,\zeta)}\sup_{f_\varepsilon\in\mathcal{D}^\beta(R,\gamma)}P\Big(\exists h\in\mathcal{B}:|\widetilde{\sigma}_{h,X}-\sigma_{h,X}|>\eta m^{-1/2}(h^{-\beta+1/2}\vee 1)\Big)\to 0.$$

Proof. Note that

$$\begin{aligned}\widetilde{\sigma}_{h,X}^2=&\frac{1}{n^2}\sum_{j=1}^n\xi_{j,1}^2(h)+\frac{1}{n^2}\sum_{j=1}^n\xi_{j,2}^2(h)+\frac{1}{n^2}\sum_{j=1}^n\xi_{j,3}^2(h)\\ &+\frac{2}{n^2}\sum_{j=1}^n\xi_{j,1}(h)\xi_{j,2}(h)+\frac{2}{n^2}\sum_{j=1}^n\xi_{j,1}(h)\xi_{j,3}(h)+\frac{2}{n^2}\sum_{j=1}^n\xi_{j,2}(h)\xi_{j,3}(h),\end{aligned}\tag{4.68}$$

where we have defined

$$\xi_{j,1}(h) := \int_{-\infty}^{0} a_s(x)\,\mathcal{F}^{-1}\Big[\varphi_K(hu)e^{iuY_j}\Big(\frac{1}{\varphi_{\varepsilon,m}(u)} - \frac{1}{\varphi_\varepsilon(u)}\Big)\Big](x+\widetilde{q}_{\tau,h})\mathrm{d}x,$$
$$\xi_{j,2}(h) := \int_{-\infty}^{0} a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x+q_\tau)\mathrm{d}x,$$
$$\xi_{j,3}(h) := \int_{-\infty}^{0} a_s(x)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}(e^{-iu\widetilde{q}_{\tau,h}} - e^{-iuq_\tau})}{\varphi_\varepsilon(u)}\Big](x)\mathrm{d}x.$$

We will first study these three terms separately. Applying Plancherel's identity, the Cauchy–Schwarz inequality, the Neumann type bound (4.33) as well as $|\mathcal{F} a_s(u)| \leqslant A_s(1+|u|)^{-1}$, the decay of φ_ε and the upper bound on f, we obtain

$$\mathbb{E}[|\xi_{j,1}(h)|^2 \mathbb{1}_{B_\varepsilon(h)}] \leqslant \frac{9}{2\pi^2}\int_{-1/h}^{1/h}\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u \int_{-1/h}^{1/h}\frac{|\varphi_K(hu)|^2}{m|\varphi_\varepsilon(u)|^2}\mathrm{d}u \lesssim \frac{1}{(h^{2\beta-1}\wedge 1)mh^{2\beta+1}}, \tag{4.69}$$

$$\mathbb{E}[|\xi_{j,2}(h)|^2] = \mathbb{E}\Big[\Big|\frac{1}{2\pi}\int \mathcal{F} a_s(u)e^{-iuq_\tau}\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}e^{iuY_j}\mathrm{d}u\Big|^2\Big] \leqslant \frac{\|K\|_{L^1}^2 A_s^2 R^3}{4\pi^2}\int_{-1/h}^{1/h}(1+|u|)^{2\beta-2}\mathrm{d}u =: S_h^2 \tag{4.70}$$

as well as the deterministic bound

$$|\xi_{j,2}(h)|^2 = \Big|\frac{1}{2\pi}\int \mathcal{F} a_s(u)e^{-iuq_\tau}\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}e^{iuY_j}\mathrm{d}u\Big|^2 \leqslant \frac{\|K\|_{L_1}^2 A_s^2}{4\pi^2}\int_{-1/h}^{1/h}(1+|u|)^{2\beta}\mathrm{d}u =: d_h^2. \tag{4.71}$$

Hence, $\mathrm{Var}[\xi_{j,2}(h)^2] \leqslant \mathbb{E}[\xi_{j,2}(h)^4] \leqslant d_h^2 S_h^2$. and $|\xi_{j,2}^2(h) - \mathbb{E}[\xi_{j,2}^2(h)]| \leqslant 2d_h^2$, so that an application of Bernstein's inequality yields for any $h > 0$ and $z > 0$

$$P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}(\xi_{j,2}^2(h) - \mathbb{E}[\xi_{j,2}^2(h)])\Big| \geqslant z\Big) \leqslant 2\exp\Big(-\frac{z^2 n}{2S_h^2 d_h^2 + \frac{4}{3}d_h^2 z}\Big).$$

Setting $z = S_h^2$ and noting $S_h^2 \lesssim (h^{-2\beta+1}\vee 1), d_h^2 \lesssim h^{-2\beta}$, we see that

$$P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}(\xi_{j,2}^2(h) - \mathbb{E}[\xi_{j,2}^2(h)])\Big| \geqslant S_h^2\Big) \leqslant 2\exp\Big(-\frac{S_h^2 n}{4d_h^2}\Big) \leqslant 2\exp\Big(-Cnh^{2\beta\wedge 1}\Big) \tag{4.72}$$

for some $C > 0$. The right-hand side of (4.72) tends to zero with polynomial rate since $nh^{2\beta\wedge 1} \gtrsim \log n$.

We use $\operatorname{supp} a_s \subseteq [-1, 0]$ to write $\xi_{j,3}$ as

$$\xi_{j,3}(h) = \int \big(a_s(x-\widetilde{q}_{\tau,h}) - a_s(x-q_\tau)\big)\,\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x)\mathrm{d}x$$
$$\leqslant \sup_{t\in(-1,0)}|a_s'(t)||\widetilde{q}_{\tau,h} - q_\tau|\int_{(\widetilde{q}_{\tau,h}\wedge q_\tau)-1}^{\widetilde{q}_{\tau,h}\vee q_\tau}\Big|\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x)\Big|\mathrm{d}x.$$

The Cauchy–Schwarz inequality and Plancherel's identity yield

$$\begin{aligned}|\xi_{j,3}(h)|^2 \leqslant &\|a_s'\mathbb{1}_{(-1,0)}\|_\infty^2|\widetilde{q}_{\tau,h}-q_\tau|^2(1+|\widetilde{q}_{\tau,h}-q_\tau|)\\ &\times\int_{(\widetilde{q}_{\tau,h}\wedge q_\tau)-1}^{\widetilde{q}_{\tau,h}\vee q_\tau}\Big|\mathcal{F}^{-1}\Big[\frac{\varphi_K(hu)e^{iuY_j}}{\varphi_\varepsilon(u)}\Big](x)\Big|^2\mathrm{d}x\\ \leqslant&\frac{\|a_s'\mathbb{1}_{(-1,0)}\|_\infty^2}{2\pi}|\widetilde{q}_{\tau,h}-q_\tau|^2(1+|\widetilde{q}_{\tau,h}-q_\tau|)\int\Big|\frac{\varphi_K(hu)}{\varphi_\varepsilon(u)}\Big|^2\mathrm{d}u\\ \lesssim&|\widetilde{q}_{\tau,h}-q_\tau|^2(1+|\widetilde{q}_{\tau,h}-q_\tau|)h^{-2\beta-1}.\end{aligned}$$

By Lemma 4.16 $\sup_{h\in\mathcal{B}}|\widetilde{q}_{\tau,h}-q_\tau| = o_P(1)$. Applying (4.66), we conclude for some constant $C>0$, for $\delta_h = (h^{\alpha+(1/2-\beta)_+}+n^{-1/2}h^{-\beta-1/2}))^{-1}$ and for any $\eta>0$

$$\begin{aligned}&P\Big(\exists h\in\mathcal{B}:|\xi_{j,3}(h)|>\eta(h^{-\beta+1/2}\vee 1)\Big)\\ \leqslant&P\Big(\exists h\in\mathcal{B}:|\widetilde{q}_{\tau,h}-q_\tau|>\eta C h^{(\beta\wedge 1/2)+1/2}\Big)+o(1)\\ \leqslant&P\Big(\exists h\in\mathcal{B}:|\widetilde{q}_{\tau,h}-q_\tau|>\eta C\delta_h(h^{\alpha+1}+n^{-1/2}(h^{-\beta+1/2}\vee 1))\Big)+o(1)\\ \lesssim&\Big(\sum_{h\in\mathcal{B}}(\delta_h)^{-1}\Big)+o(1)\lesssim\sup_{h\in\mathcal{B}}h^\alpha\log n+\sup_{h\in\mathcal{B}}\frac{\log n}{\sqrt{n}h^{\beta+1/2}}+o(1)=o(1).\end{aligned}\tag{4.73}$$

Combining the variance bounds (4.69), (4.70) and (4.73), we apply Markov's inequality, the Cauchy–Schwarz inequality and the concentration result (4.72) on the decomposition (4.68) to obtain

$$\begin{aligned}\sup_{h\in\mathcal{B}}\Big(n(h^{2\beta-1}\wedge 1)|\widetilde{\sigma}_{h,X}^2-\sigma_{h,X}^2|\Big)&=\sup_{h\in\mathcal{B}}\Big(\frac{h^{2\beta-1}\wedge 1}{n}\sum_{j=1}^n(\xi_{j,2}^2(h)-\mathbb{E}[\xi_{j,2}^2(h)])\Big)+o_P(1)\\ &=o_P(1).\end{aligned}$$

□

Lemma 4.20. *Let $\mathcal{B}$ be a finite set satisfying $|\mathcal{B}|\lesssim\log n$ and $\sup_{h\in\mathcal{B}}1/(nh^{2\beta+1})\to 0$. Let $\widetilde{\sigma}_{h,\varepsilon}$ and $\sigma_{h,\varepsilon}$ be given in (4.17) and* (4.62) *respectively. Then we obtain uniformly over $f\in\mathcal{C}^\alpha(R,r,\zeta)$ and $f_\varepsilon\in\mathcal{D}^\beta(R,\gamma)$ for all $\eta>0$ as $n\to\infty$*

$$P\Big(\exists h\in\mathcal{B}:|\widetilde{\sigma}_{h,\varepsilon}-\sigma_{h,\varepsilon}|>\eta(\log n)m^{-1/2}(h^{-\beta+1/2}\vee 1)\Big)\to 0.$$

Proof. We start by showing for $h_1=\min\mathcal{B}$ that

$$\sup_{|u|\leqslant 1/h_1}\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}\Big|=1+o_P(1).\tag{4.74}$$

To this end, recall $w(u)=(\log(e+|u|))^{-1/2-\eta}$ for some $\eta\in(0,1/2)$. Markov's inequality, Lemma 4.13 and Theorem 4.1 by Neumann and Reiß (2009) yield for any $\delta > 0$

$$
\begin{aligned}
P\Big(\sup_{|u|\leqslant 1/h_1}\Big|\frac{\varphi_\varepsilon(u)}{\varphi_{\varepsilon,m}(u)}-1\Big|\geqslant\delta\Big) \leqslant & P\Big(\sup_{|u|\leqslant 1/h_1} m^{1/2}|\varphi_\varepsilon(u)-\varphi_{\varepsilon,m}(u)|\geqslant\delta|\log h_1|\Big)\\
&+P\Big(\inf_{|u|\leqslant 1/h_1}|\varphi_{\varepsilon,m}(u)|\leqslant m^{-1/2}|\log h_1|\Big)\\
\leqslant & (\delta|\log h_1|)^{-1}\,\mathbb{E}\Big[\sup_{|u|\leqslant 1/h_1} m^{1/2}|\varphi_\varepsilon(u)-\varphi_{\varepsilon,m}(u)|\Big]+o(1)\\
\leqslant & \frac{m^{1/2}}{\delta|\log h_1|w(1/h_1)}\,\mathbb{E}\Big[\sup_{u\in\mathbb{R}} w(u)|\varphi_\varepsilon(u)-\varphi_{\varepsilon,m}(u)|\Big]+o(1)\\
= & o(1),
\end{aligned}
$$

which implies (4.74). Since $[-1/h_1, 1/h_1]$ is the maximal interval for all $h \in \mathcal{B}$, (4.74) holds uniformly in $\mathcal{B}$.

Now, we consider $\widetilde{\sigma}_{h,\varepsilon,1}$. The uniform consistency (4.74) implies

$$
\widetilde{\sigma}^2_{h,\varepsilon,1} = (1+o_P(1))\int|\varphi_K(hu)|\Big|\frac{\varphi_n(u)}{\varphi_\varepsilon(u)}\Big|^2\mathrm{d}u.
$$

Chebyshev's inequality yields for all $\eta > 0$

$$
\begin{aligned}
&P\Big(\sup_{h\in\mathcal{B}}\Big|\Big(\int|\varphi_K(hu)|\frac{|\varphi_n(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)^{1/2}-\mathbb{E}\Big[\Big(\int|\varphi_K(hu)|\frac{|\varphi_n(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)^{1/2}\Big]\Big|>\eta\log n\Big)\\
&\leqslant(\eta\log n)^{-2}\sum_{h\in\mathcal{B}}\mathbb{E}\Big[\int|\varphi_K(hu)|\frac{|\varphi_n(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big]\\
&\lesssim(\eta^2\log n)^{-1}\int_{-1/h_1}^{1/h_1}\frac{\mathbb{E}[|\varphi_n(u)|^2]}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\lesssim(\eta^2\log n)^{-1},
\end{aligned}
$$

where the last estimate follows from $\mathbb{E}[|\varphi_n(u)|^2] \lesssim |\varphi_Y(u)|^2 + \mathbb{E}[|\varphi_n(u) - \varphi_Y(u)|^2] \lesssim |\varphi_Y(u)|^2 + 1/n$, $f_\varepsilon \in \mathcal{D}^\beta(R, \gamma)$, $\|f\|_\infty \lesssim 1$ and $nh_1^{2\beta+1} \to \infty$. Hence, we obtain uniformly in $\mathcal{B}$

$$
\widetilde{\sigma}_{h,\varepsilon,1} = (1+o_P(1))\big(\sigma_{h,\varepsilon,1}+o_P(\log n)\big) = \sigma_{h,\varepsilon,1}+o_P(\log n). \tag{4.75}
$$

Concerning $\widetilde{\sigma}_{h,\varepsilon,2}$, we write with use of (4.74)

$$
\widetilde{\sigma}^2_{h,\varepsilon,2} = \int|\varphi_K(hu)|\frac{|\mathcal{F}a_s(u)|^2}{|\varphi_{\varepsilon,m}(u)|^2}\mathrm{d}u = (1+o_p(1))\int|\varphi_K(hu)|\frac{|\mathcal{F}a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u.
$$

Moreover, the triangle inequality for the L^2-norm and Lemma 4.13, applied on $B_\varepsilon(h_1)$

yield

$$
\begin{aligned}
&\Big|\Big(\int |\varphi_K(hu)|\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)^{1/2} - \sigma_{h,\varepsilon,2}\Big|^2\\
\leqslant& 2\Big|\,\mathbb{E}\Big[\Big(\Big(\int |\varphi_K(hu)|\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)^{1/2} - \Big(\int |\varphi_K(hu)|\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_{\varepsilon,m}(u)|^2}\mathrm{d}u\Big)^{1/2}\Big)\mathbb{1}_{B_\varepsilon(h_1)}\Big]\Big|^2\\
&\quad + 2P\big((B_\varepsilon(h_1))^c\big)\int |\varphi_K(hu)|\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\\
\leqslant& 2\,\mathbb{E}\Big[\Big(\int |\varphi_K(hu)||\mathcal{F} a_s(u)|^2\frac{|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^2}{|\varphi_\varepsilon(u)\varphi_{\varepsilon,m}(u)|^2}\mathrm{d}u\Big)\mathbb{1}_{B_\varepsilon(h_1)}\Big]\\
&\quad + o(1)\int_{-1/h}^{1/h}\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\\
\leqslant& \frac{2}{|\log h_1|^{3/2}}\,\mathbb{E}\Big[\int_{-1/h}^{1/h}\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}m|\varphi_{\varepsilon,m}(u)-\varphi_\varepsilon(u)|^2\mathrm{d}u\Big] + o(1)(h^{-2\beta+1}\vee 1)\\
=& o(1)(h^{-2\beta+1}\vee 1),
\end{aligned}
$$

where $o(1)$ is a null sequence which does not depend on h. Consequently,

$$\sup_{h\in\mathcal{B}}\Big|\Big(\int |\varphi_K(hu)|\frac{|\mathcal{F} a_s(u)|^2}{|\varphi_\varepsilon(u)|^2}\mathrm{d}u\Big)^{1/2} - \sigma_{h,\varepsilon,2}\Big|(h^{\beta-1/2}\wedge 1) = o(1).$$

Using $\sigma_{h,\varepsilon,2}^2 \lesssim h^{-2\beta+1}\vee 1$ by the analysis of the convergence rates, we get

$$\widetilde{\sigma}_{h,\varepsilon,2} = (1+o_p(1))\big(\sigma_{h,\varepsilon,2} + o\big(h^{-\beta+1/2}\vee 1\big)\big) = \sigma_{h,\varepsilon,2} + o_P\big(h^{-\beta+1/2}\vee 1\big). \tag{4.76}$$

Since $\sigma_{h,\varepsilon,1}\lesssim 1, \sigma_{h,\varepsilon,2}\lesssim h^{-\beta+1/2}\vee 1$, it remains to combine (4.75) and (4.76) to obtain uniformly in $\mathcal{B}$

$$
\begin{aligned}
\widetilde{\sigma}_{h,\varepsilon} &= \frac{1}{2\pi}m^{-1/2}\widetilde{\sigma}_{h,\varepsilon,1}\widetilde{\sigma}_{h,\varepsilon,2} = \frac{1}{2\pi}m^{-1/2}\big(\sigma_{h,\varepsilon,1} + o_P(\log n)\big)\big(\sigma_{h,\varepsilon,2} + o_P\big(h^{-\beta+1/2}\vee 1\big)\big)\\
&= \sigma_{h,\varepsilon} + o_P\big((\log n)m^{-1/2}(h^{-\beta+1/2}\vee 1)\big).
\end{aligned}
$$

□

Proof of Theorem 4.12

Applying Lemma 4.13 and (4.53), it suffices to consider the event

$$A_0 := \{h_{j_{0,n}} \leqslant h_{\widetilde{j}_n} \leqslant n^{-1/(2\alpha+2(\beta\vee 1/2)+1)}\} \cap B_\varepsilon(h_{j_{0,n}})$$

with $j_{0,n}$ defined in (4.52). Therefore we can set $\mathcal{B} := \{h_{j_{0,n}},\dots,h_{M_n}\}$ in the following.

As seen in error decomposition (4.56), there are three stochastic errors $V_{h,X}, V_{h,\varepsilon}$ and $V_{h,c}$ which were treated in Lemma 4.17. This motivates the following definition. For $\delta_1 > 0$ let

$$S_{h,X} := (1+\delta_1)\sqrt{2\log\log n}\max_{\mu\in\mathcal{B}:\mu\geqslant h}\sigma_{\mu,X},\quad S_{h,\varepsilon} := (\delta_1\log n)^3\max_{\mu\in\mathcal{B}:\mu\geqslant h}\sigma_{\mu,\varepsilon}.$$

On the assumption $|\varphi_\varepsilon(u)| \gtrsim (1+|u|)^{-\beta}$ we obtain for $\sigma_{h,\varepsilon} = \frac{1}{2\pi} m^{-1/2}\sigma_{h,\varepsilon,1}\sigma_{h,\varepsilon,2}$ from (4.62) that

$$\sigma_{h,\varepsilon,2}^2 \gtrsim \int_{-1/h}^{1/h} |\mathcal{F} a_s(-u)|^2 (1+|u|)^{2\beta} \mathrm{d}u \gtrsim \int_{-1/h}^{1/h} (1+|u|)^{2\beta-2} \mathrm{d}u \sim h^{-2\beta+1} \vee 1.$$

Also, we have $\sigma_{h,\varepsilon,1} = \| \varphi_X \|_{L^2} + o(1) \geq \| \varphi_X \|_{L^2} /2$ for h small enough and n large enough. Thus $\sigma_{h,\varepsilon} \gtrsim m^{-1/2}(h^{-\beta+1/2} \vee 1)$. Therefore, Lemma 4.17 yields

$$\begin{aligned}
&P\Big(\exists h \in \mathcal{B} : |V_{h,X} + V_{h,\varepsilon} + V_{h,c}| \geqslant S_{h,X} + S_{h,\varepsilon}\Big) \\
&\quad \leqslant P\Big(\exists h \in \mathcal{B} : |V_{h,X}| \geqslant S_{h,X} + \tfrac{1}{3} S_{h,\varepsilon}\Big) + P\Big(\exists h \in \mathcal{B} : |V_{h,\varepsilon}| \geqslant \frac{S_{h,\varepsilon}}{3}\Big) \\
&\qquad + P\Big(\exists h \in \mathcal{B} : |V_{h,c}| \geqslant \frac{S_{h,\varepsilon}}{3}\Big) \\
&\quad = o(1).
\end{aligned}$$

Hence, the probability of the event

$$A_1 := \Big\{\forall h \in \mathcal{B} : |V_{h,X} + V_{h,\varepsilon} + V_{h,c}| \leqslant S_{h,X} + S_{h,\varepsilon}\Big\}$$

converges to one. The variances $S_{h,X}$ and $S_{h,\varepsilon}$ can be estimated by

$$\widetilde{S}_{h,X} := (1+\delta_1)\sqrt{2 \log\log n} \max_{\mu \in \mathcal{B}: \mu \geqslant h} \widetilde{\sigma}_{\mu,X}, \quad \widetilde{S}_{h,\varepsilon} := (\delta_1 \log n)^3 \max_{\mu \in \mathcal{B}: \mu \geqslant h} \widetilde{\sigma}_{\mu,\varepsilon}.$$

Applying Lemmas 4.19 and 4.20, the triangle inequality of the ℓ^∞-norm yields uniformly in $h \in \mathcal{B}$

$$\begin{aligned}
|\max_{\mu \geqslant h} \widetilde{\sigma}_{\mu,X} - \max_{\mu \geqslant h} \sigma_{\mu,X}| &\leqslant \max_{\mu \geqslant h} |\widetilde{\sigma}_{\mu,X} - \sigma_{\mu,X}| = o_P\Big(\frac{1}{m^{1/2}(h^{\beta-1/2} \wedge 1)}\Big), \\
|\max_{\mu \geqslant h} \widetilde{\sigma}_{\mu,\varepsilon} - \max_{\mu \geqslant h} \sigma_{\mu,\varepsilon}| &\leqslant \max_{\mu \geqslant h} |\widetilde{\sigma}_{\mu,\varepsilon} - \sigma_{\mu,\varepsilon}| = o_P\Big(\frac{\log n}{m^{1/2}(h^{\beta-1/2} \wedge 1)}\Big).
\end{aligned}$$

Using again $\sigma_{h,\varepsilon} \gtrsim m^{-1/2}(h^{-\beta+1/2} \vee 1)$, we thus obtain for all $\eta > 0$ that the event

$$A_2 := \Big\{\forall h \in \mathcal{B} : \big|(\widetilde{S}_{h,X} + \widetilde{S}_{h,\varepsilon}) - (S_{h,X} + S_{h,\varepsilon})\big| \leqslant \eta (S_{h,X} + S_{h,\varepsilon})\Big\}$$

fulfills $P(A_2) \to 1$. The same holds true for the events

$$\begin{aligned}
A_3 := &\Big\{\forall h \in \mathcal{B} : \sup_{q^* \in [(q_\tau \wedge \widetilde{q}_{\tau,h}) \vee (q_\tau \wedge \widetilde{q}_{\tau,h})]} |\widetilde{f}_h(q^*) - f(q_\tau)| \leqslant \eta f(q_\tau)\Big\}, \\
A_4 := &\Big\{\forall h \in \mathcal{B} : \sup_{q^* \in [(q_\tau \wedge \widetilde{q}_{\tau,h}) \vee (q_\tau \wedge \widetilde{q}_{\tau,h})]} |\widetilde{f}_h(q^*) - \widetilde{f}_h(\widetilde{q}_{\tau,h})| \leqslant \eta |\widetilde{f}_h(\widetilde{q}_{\tau,h})|\Big\}
\end{aligned}$$

by (4.65). Denote the event from (4.39) by A_5 Therefore, it is sufficient to work in the following on the event

$$A := A_0 \cap A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5.$$

We show that the adaptive estimator $\widetilde{q}_\tau$ mimics the oracle estimator defined as follows. Recalling the estimate of the bias $B_h = Dh^{\alpha+1}$, let the oracle bandwidth be defined by

$$h_* := \max\{h \in \mathcal{B} : B_h \leqslant S_{h,X} + S_{h,\varepsilon}\}. \tag{4.77}$$

Note that h_* is well-defined and unique since B_h is monoton increasing in h while $(S_{h,X} + S_{h,\varepsilon})$ is monton decreasing. We get the oracle estimator $\widetilde{q}_{\tau,h_*}$.

Since on A_4 for all $h \in \mathcal{B}$ and $q^* \in [(q_\tau \wedge \widetilde{q}_{\tau,h}) \vee (q_\tau \wedge \widetilde{q}_{\tau,h})]$

$$|\widetilde{f}_h(q^*)| \geqslant |\widetilde{f}_h(\widetilde{q}_{\tau,h})| - |\widetilde{f}_h(q^*) - \widetilde{f}_h(\widetilde{q}_{\tau,h})| \geqslant (1-\eta)|\widetilde{f}_h(\widetilde{q}_{\tau,h})|,$$

we have for any $h \in \mathcal{B}$ on the event $A_1 \cap A_4 \cap A_5$ by (4.56) and (4.38)

$$|\widetilde{q}_{\tau,h} - q_\tau| \leqslant \frac{B_h + |V_{h,X} + V_{h,\varepsilon} + V_{h,c}|}{|\widetilde{f}_h(q^*)|} \leqslant \frac{B_h + S_{h,X} + S_{h,\varepsilon}}{(1-\eta)|\widetilde{f}_h(\widetilde{q}_{\tau,h})|}.$$

Furthermore, by the definition of h_* we have on the event A for any $h \leqslant h_*$

$$|\widetilde{q}_{\tau,h} - q_\tau| \leqslant \frac{2(S_{h,X} + S_{h,\varepsilon})}{(1-\eta)|\widetilde{f}_h(\widetilde{q}_{\tau,h})|}.$$

On A_2 we estimate $\widetilde{S}_{h,X} + \widetilde{S}_{h,\varepsilon} \geqslant (1-\eta)(S_{h,X} + S_{h,\varepsilon})$ and thus we have on A for any $h \leqslant h_*$

$$|\widetilde{q}_{\tau,h} - q_\tau| \leqslant \frac{2(\widetilde{S}_{h,X} + \widetilde{S}_{h,\varepsilon})}{(1-\eta)^2|\widetilde{f}_h(\widetilde{q}_{\tau,h})|}.$$

Since for any $\delta > 0$ we find $\delta_1, \eta > 0$ such that $\big((1-\eta)^{-2}(2\sqrt{2}+\delta_1) - 2\sqrt{2}\big) \vee (2(1-\eta)^{-2}\delta_1) < \delta$, we obtain $|\widetilde{q}_{\tau,h} - q_\tau| \leqslant \widetilde{\Sigma}_h$ with $\widetilde{\Sigma}_h$ as defined in (4.15). As a result one has $q_\tau \in \mathcal{U}_h$ and $q_\tau \in \mathcal{U}_\mu$ for all $h \leqslant h_*$ and $\mu \leqslant h_*$, implying $\mathcal{U}_\mu \cap \mathcal{U}_h \neq \emptyset$. By the definition of the procedure, $\widetilde{h}^* \geqslant h_*$ and $\mathcal{U}_{\widetilde{h}_*} \cap \mathcal{U}_{h_*} \neq \emptyset$ on the event A. This leads to

$$|\widetilde{q}_{\tau,\widetilde{h}^*} - q_\tau| \leqslant |\widetilde{q}_{\tau,h_*} - q_\tau| + |\widetilde{q}_{\tau,\widetilde{h}^*} - \widetilde{q}_{\tau,h_*}| \leqslant \widetilde{\Sigma}_{h_*} + (\widetilde{\Sigma}_{h_*} + \widetilde{\Sigma}_{\widetilde{h}^*})$$

On $A_2 \cap A_3$ we have $\widetilde{\Sigma}_h \lesssim S_{h,X} + S_{h,\varepsilon}$ since $f(q_\tau) \geqslant r$. Using additionally the monotonicity of $(S_{h,X} + S_{h,\varepsilon})$ as well as $\widetilde{h}^* \geqslant h_*$, this implies

$$|\widetilde{q}_{\tau,\widetilde{h}^*} - q_\tau| \lesssim (S_{h_*,X} + S_{h_*,\varepsilon}) \lesssim \Big(\sqrt{\log\log n} + \big(\log n^\delta\big)^3\Big)\big(h_*^{-\beta+1/2} \vee 1\big)n^{-1/2}.$$

It remains to note by the definition (4.77) of the oracle h_* and by the assumption $h_{j+1}/h_j \lesssim 1$ that $h_* \sim ((\log n^\delta)^6/n)^{-1/(2\alpha+2(\beta\vee 1/2)+1)}$ as $n \to \infty$. □

5. Quantile estimation for Lévy measures

After we have studied in full detail quantile estimation in the deconvolution framework in Chapter 4, we consider in this chapter the more involved problem of estimating the generalized quantiles of a Lévy process. Compared to deconvolution, the Lévy model is harder for two reasons. First, it is a nonlinear inverse problem such that we have to linearize the estimation error and the remainder needs extra care. Second, as we saw in Chapter 3 the underlying deconvolution problem is determined by the distribution of the process itself. Consequently, there is a strong interplay between the jump measure that we want to estimate and the corresponding deconvolution operator.

Let $L = \{L_t : t \geqslant 0\}$ be a real-valued Lévy process with characteristic triplet (σ^2, γ, ν). Taking into account the possible singularity of ν at zero, Nickl and Reiß (2012) as well as Nickl et al. (2013) have considered the estimation problem of the (generalized) distribution function

$$N(t) := \begin{cases} \nu((-\infty, t]), & \text{for } t < 0, \\ \nu([t, \infty)), & \text{for } t > 0, \end{cases}$$

when it can be estimated with the parametric rate. We aim for its inverse function in a more general situation. Assuming absolute continuity of ν at $\pm q_\tau^\pm$ with respect to the Lebesgue measure, the generalized quantiles $q_\tau^+ > 0$ and $q_\tau^- > 0$ for a given level $\tau \in (0, \nu(\mathbb{R}_\pm))$, introduced in Chapter 1, are determined by

$$N(-q_\tau^-) = \tau = N(q_\tau^+).$$

If ν has finite mass on the negative or the positive halfline, the τ-quantiles only exist if $\tau \leqslant \nu(\mathbb{R}_-)$ or $\tau \leqslant \nu(\mathbb{R}_+)$, respectively. Since ν is not known, it is reasonable to estimate $q_\tau^\pm \vee \eta_n$ for some threshold $\eta_n > 0$ and any $\tau > 0$ instead of the quantiles themselves. As $n \to \infty$ the threshold value η_n may converge slowly to zero. Then the estimators have the interpretation that if η_n is attained then probably $\tau > \nu(\mathbb{R}_\pm)$.

Suppose we have access to an estimator of the characteristic function φ_t of the marginal distribution L_t of the Lévy process for some $t > 0$. Assuming $\int x^2 \nu(\mathrm{d}x) < \infty$, φ_t is given by the Lévy–Khintchine representation in Kolmogorov's version, cf. Proposition 2.3,

$$\varphi_t(u) := \mathbb{E}[e^{iuY_t}] = e^{t\psi(u)} \quad \text{with} \quad \psi(u) := -\frac{\sigma^2}{2}u^2 + i\gamma_1 u + \int (e^{iux} - 1 - iux)\nu(\mathrm{d}x)$$

with $\gamma_1 = \gamma - \int_{|x|>1} x\nu(\mathrm{d}x)$. Taking the second derivative of the characteristic exponent ψ, we obtain the estimating equation

$$\psi''(u) = -\sigma^2 - \mathcal{F}[x^2\nu](u) = \frac{\varphi_t''(u)\varphi_t(u) - \varphi_t'(u)^2}{t\varphi_t^2(u)}. \tag{5.1}$$

To estimate $q_\tau^\pm$, we will follow the program that we have outlined in the introduction:

(i) A density estimator for the jump measure ν can be constructed by replacing φ_t in (5.1) with its estimator and regularizing with a band limited kernel.

(ii) The plug-in approach yields an estimator for the distribution function.

(iii) The generalized τ-quantiles can be estimated by minimizing the distance between the value of distribution function estimator and τ.

Before we state these considerations more precisely we will see in Section 5.1 that the Fourier multiplier property of the deconvolution operator, which is essential in the mathematical analysis in Chapter 4, is naturally satisfied for infinitely divisible distributions. Afterwards, we study two different observations schemes. In Section 5.2 equidistant discrete observations of the Lévy process are considered where φ_t can be estimated by the empirical measure of the increments. In Section 5.3 we revisit the financial example from Chapter 2. Observing option prices, the characteristic function can be estimated via the pricing formula that links φ_t and the option function. All proofs concerning the quantile estimation are postponed to Section 5.4.

5.1. A Fourier multiplier theorem for infinitely divisible distributions

Let μ be an infinitely divisible distribution with characteristic triplet (σ^2, γ, ν). Denote the characteristic function of μ by

$$\varphi(u) := \int e^{iux}\mu(\mathrm{d}x) = \exp\Big(-\frac{\sigma^2}{2}u^2 + i\gamma u + \int (e^{iux} - 1 - iux\mathbb{1}_{[-1,1]}(x))\nu(\mathrm{d}x)\Big),$$

where the latter equality is given by the Lévy-Khintchine formula. The Fourier multiplier $1/\varphi$ which gives rise to the map $f \mapsto \mathcal{F}^{-1}[\mathcal{F} f/\varphi]$ for appropriate functions f has a prominent role in the Chapters 3 and 4. Applying a Mihlin multiplier theorem, Nickl and Reiß (2012) quantified the mapping properties of the Fourier multiplier in a Besov scale under certain assumptions. The target of this section is to study necessary and sufficient conditions under which $1/\varphi$ is a Fourier multiplier on Besov spaces or equivalently there is some $\alpha \geqslant 0$ such that for any $s \in \mathbb{R}, p, q \in [1, \infty]$ the deconvolution map

$$B^{s+\alpha}_{p,q}(\mathbb{R}) \ni f \mapsto \mathcal{F}^{-1}[\mathcal{F} f/\varphi] \in B^s_{p,q}(\mathbb{R})$$

is bounded. we again use the notation $\mathcal{F}^{-1}[1/\varphi]*f := \mathcal{F}^{-1}[\mathcal{F} f/\varphi]$ and refer to $\mathcal{F}^{-1}[1/\varphi]$ as the deconvolution operator.

Using that $B^s_{p,q}(\mathbb{R}) \ni f \mapsto \mathcal{F}^{-1}[(1+iu)^\alpha \mathcal{F} f] \in B^{s+\alpha}_{p,q}(\mathbb{R})$ is an isomorphism for any $\alpha \in \mathbb{R}$ and that the set of Fourier multipliers on $B^s_{p,q}(\mathbb{R})$ does not depend on s, q and are nested in p (Triebel, 2010, Prop. 2.6.2 and Thm. 2.6.3), the characteristic function φ has to satisfy necessarily

$$|\varphi(u)| \gtrsim (1 + |u|)^{-\alpha}, \quad u \in \mathbb{R}.$$

Our first aim is therefore to characterize infinitely divisible distributions whose characteristic function decays only with a polynomial rate.

Due to the connection between the decay of the characteristic exponent and the regularity of transition densities established by Hartman and Wintner (1942), upper bounds for $|\varphi|$ have attracted a lot interest in the literature. Orey (1968); Kallenberg (1981); Knopova and Schilling (2013) study necessary and sufficient conditions for an exponential decay of φ and thus for the existence of infinitely smooth transition densities. For self-decomposable distributions the existence of polynomial upper bounds is in detail analyzed, see Sato (1999, Chap. 28) and references therein. While upper bounds for $|\varphi|$ are more interesting from the probabilistic perspective, lower bounds are highly relevant from a statistical point of view. In particular, a polynomially decaying characteristic function corresponds to mildly ill-posed estimation problems. In Trabs (2014) a polynomial lower bound is established for a class of Lévy processes which is closely related to self-decomposable processes.

Defining the symmetrized jump measure by

$$\nu_s(A) := \nu(A) + \nu(-A)$$

for any Borel set $A \in \mathscr{B}(\mathbb{R}_+)$, the absolute value of the characteristic function is given by

$$|\varphi(u)| = \exp\Big(-\frac{\sigma^2 u^2}{2} + \int (\cos(ux) - 1)\nu(\mathrm{d}x)\Big) = \exp\Big(-\frac{\sigma^2 u^2}{2} + \int_0^\infty (\cos(ux) - 1)\nu_s(\mathrm{d}x)\Big).$$

The following result asserts that under a mild regularity assumption on ν in a neighborhood of the origin, φ decays polynomially if and only if there is no diffusion component and the Lebesgue density of $x\nu(\mathrm{d}x)$ (which we assume to exist at zero) is bounded. In the appendix we define the space of bounded variation functions $BV(\mathbb{R})$. Recall that for any $f \in BV(\mathbb{R})$ there is a finite signed Borel measure Df such that $Df((a,b]) = f(b+) - f(a+)$ for $-\infty < a < b < \infty$ and $\|f\|_{BV} = \|Df\|_{TV}$ with the total variation norm $\|\bullet\|_{TV}$.

Assumption 5.A. Assume that $x\nu_s(\mathrm{d}x)$ admits on a small interval $(0,\delta]$, for some $\delta \in (0,1)$, a Lebesgue density k_s with $k_s \in BV([\varepsilon,\delta])$ for all $\varepsilon \in (0,\delta)$. Suppose that $\|(Dk_s)^+|_{(0,\delta]}\|_{TV} = \lim_{\varepsilon\downarrow 0} \|(Dk_s)^+|_{[\varepsilon,\delta]}\|_{TV} < \infty$.

Assumption 5.A basically excludes Lévy measures which oscillate at zero or have additional singularities in any neighborhood of the origin. Both possibilities are not natural in applications, for instance, for the modeling of stochastic processes via Lévy processes.

Proposition 5.1. *Let μ be an infinitely divisible distribution which satisfies Assumption 5.A. Then the following are equivalent:*

(i) $\sigma^2 = 0$ *and* $\|k_s \mathbb{1}_{(0,\delta]}\|_{BV} < \infty$,

(ii) $\sigma^2 = 0$ *and* $\lim_{x\to 0} k_s(x) < \infty$,

(iii) there is some $\alpha > 0$ and some constant $c > 0$ such that $|\varphi(u)| \geqslant c(1 + |u|)^{-\alpha}$ for all $u \in \mathbb{R}$.

This proposition is inspired by the behavior of self-decomposable distributions which we have introduced in Chapter 2. To prove it, we will generalize Lemma 2.1 by Trabs (2014) and its counterpart Lemma 53.9 in Sato (1999). The following lemma shows that a polynomial decay of the characteristic function holds true for a class of infinitely divisible distributions which is much larger than self-decomposable distributions.

Lemma 5.2. *If $\sigma^2 = 0$ and $x\nu_s(\mathrm{d}x)$ admits on an interval $(0,\delta]$, for some $\delta > 0$, a Lebesgue density $k_s\mathbb{1}_{(0,\delta]} \in BV(\mathbb{R})$, then for any $\varepsilon > 0$*

$$(1+|u|)^{-\alpha-\varepsilon} \lesssim |\varphi(u)| \lesssim (1+|u|)^{-\alpha+\varepsilon} \qquad \text{with} \quad \alpha := k_s(0+).$$

If, moreover, $\int_0^\delta \log(y^{-1})(Dk_s)^+(\mathrm{d}y) < \infty$, then $|\varphi(u)| \gtrsim (1+|u|)^{-\alpha}$.

Proof. Using

$$|\varphi(u)| = \exp\Big(\int_0^\infty \frac{\cos(ux)-1}{x} k_s(x)\mathrm{d}x\Big)$$

and the symmetry of the cosine function, we can assume $u > 0$ without loss of generality. We denote $\widetilde{\rho} = Dk_s$ and let $\tau \in (0, \delta \wedge 1)$. Since $\|k_s\mathbb{1}_{[0,\delta]}\|_\infty \leqslant \|k_s\mathbb{1}_{[0,\delta]}\|_{BV} < \infty$, we estimate for $u \leqslant 1/\tau$

$$\begin{aligned} 1 \geqslant |\varphi(u)| &= \exp\Big(\int_0^\delta \frac{\cos(ux)-1}{x} k_s(x)\mathrm{d}x + \int_\delta^\infty (\cos(ux)-1)\nu_s(\mathrm{d}x)\Big) \\ &\geqslant \exp\Big(\|k_s\|_{L^\infty([0,\delta])} \int_0^{u\delta} \frac{\cos x - 1}{x}\mathrm{d}x - 2\int_\delta^\infty \nu_s(\mathrm{d}x)\Big) \\ &\geqslant \exp\Big(-\|k_s\|_{L^\infty([0,\delta])} \sup_{v\in(0,\delta/\tau]} \int_0^v \frac{1-\cos x}{x}\mathrm{d}x - 2\int_\delta^\infty \nu_s(\mathrm{d}x)\Big), \end{aligned}$$

where the last line is a positive constant independent of u. It remains to consider the tail behavior of $|\varphi(u)|$ for $u > 1/\tau$.

To show the upper bound of $|\varphi(u)|$ for $u > 1/\tau$. We use Fubini's theorem and the finite constant $c_1 := \sup_{v\geqslant 1} \int_1^v \frac{\cos x}{x}\mathrm{d}x > 0$ to estimate

$$\begin{aligned} |\varphi(u)| &\leqslant \exp\Big(\int_{1/u}^\tau \frac{\cos(ux)-1}{x} k_s(x)\mathrm{d}x\Big) \\ &= \exp\Big(\int_{1/u}^\tau \frac{\cos(ux)-1}{x} \int_{1/u}^x \widetilde{\rho}(\mathrm{d}y)\mathrm{d}x + k_s(\tfrac{1}{u}+) \int_1^{\tau u} \frac{\cos(x)-1}{x}\mathrm{d}x\Big) \\ &= \exp\Big(\int_{1/u}^\tau \int_y^\tau \frac{\cos(ux)-1}{x}\mathrm{d}x\widetilde{\rho}(\mathrm{d}y) + k_s(\tfrac{1}{u}+)\Big(\int_1^{\tau u} \frac{\cos x}{x}\mathrm{d}x - \log(\tau u)\Big)\Big) \\ &\leqslant \exp\Big(2\int_{1/u}^\tau \big(\log\tau + \log(\tfrac{1}{y})\big)\widetilde{\rho}^-(\mathrm{d}y) - (\log u)k_s(\tfrac{1}{u}+) + (c_1 + \log(\tfrac{1}{\tau}))k_s(\tfrac{1}{u}+)\Big) \\ &\leqslant \exp\Big(-(\log u)\big(k_s(\tfrac{1}{u}+) - 2\int_{1/u}^\tau \widetilde{\rho}^-(\mathrm{d}y)\big) + (c_1 + \log(\tfrac{1}{\tau}))k_s(\tfrac{1}{u}+)\Big). \end{aligned} \tag{5.2}$$

Hence, for any $\epsilon > 0$ we find some $\tau \in (0,\delta)$ which is sufficiently small such that $|\varphi(u)| \lesssim u^{-(\alpha-\epsilon)}$ for all $u > 1/\tau$.

To verify the lower bound, we decompose for $\tau \in (0,\delta)$ and $u > 1/\tau$

$$|\varphi(u)| = \exp\Big(\Big(\int_0^{1/u} + \int_{1/u}^\tau\Big)\frac{\cos(ux)-1}{x}k_s(x)\mathrm{d}x + \int_\tau^\infty (\cos(ux)-1)\nu_s(\mathrm{d}x)\Big), \tag{5.3}$$

where the three integrals will be bounded separately from below. Using $\|k_s\mathbb{1}_{[0,\delta]}\|_\infty < \infty$, we estimate the first integral by

$$\int_0^{1/u} \frac{\cos(ux)-1}{x} k_s(x)\mathrm{d}x \geqslant \|k_s\mathbb{1}_{[0,\delta]}\|_\infty \int_0^1 \frac{\cos x - 1}{x}\mathrm{d}x,$$

where the integral on the right-hand side is a negative finite constant independent of u. The third integral in (5.3) can be bounded by

$$\int_\tau^\infty (\cos(ux)-1)\nu_s(\mathrm{d}x) \geqslant -2\int_\tau^\infty \nu_s(\mathrm{d}x).$$

It remains to estimate the second integral, where we proceed similarly to the upper bound. We obtain with the finite constant $c_2 := \inf_{v\geqslant 1}\int_1^v \frac{\cos x}{x}\mathrm{d}x < 0$

$$\begin{aligned}
&\int_{1/u}^\tau \frac{\cos(ux)-1}{x} k_s(x)\mathrm{d}x \\
&\quad = \int_{1/u}^\tau \int_y^\tau \frac{\cos(ux)-1}{x}\mathrm{d}x\widetilde{\rho}(\mathrm{d}y) + k_s(\tfrac{1}{u}+)\Big(\int_1^{\tau u} \frac{\cos x}{x}\mathrm{d}x - \log(\tau u)\Big) \\
&\quad \geqslant -2\int_{1/u}^\tau \big(\log\tau + \log(y^{-1})\big)\widetilde{\rho}^+(\mathrm{d}y) - k_s(\tfrac{1}{u}+)\log u + c_2 k_s(\tfrac{1}{u}+) \\
&\quad = -(\log u)\Big(k_s(\tfrac{1}{u}+) + 2\int_{1/u}^\tau \widetilde{\rho}^+(\mathrm{d}y)\Big) + c_2 k_s(\tfrac{1}{u}+).
\end{aligned}$$

This yields $|\varphi(u)| \gtrsim u^{-(\alpha+\epsilon)}$ for any $\epsilon > 0$ and for all $u > 1/\tau$ with τ sufficiently small.

The addendum follows from the previous estimates, the additional assumption and

$$\begin{aligned}
&\int_{1/u}^\tau \frac{\cos(ux)-1}{x} k_s(x)\mathrm{d}x \\
&\quad = \int_0^\tau \int_y^\tau \frac{\cos(ux)-1}{x}\mathrm{d}x\widetilde{\rho}(\mathrm{d}y) + k_s(0+)\Big(\int_1^{\tau u} \frac{\cos x}{x}\mathrm{d}x - \log(\tau u)\Big) \\
&\quad \geqslant -2\int_0^\tau \log(y^{-1})\widetilde{\rho}^+(\mathrm{d}y) - \alpha\log u + c_2\alpha.
\end{aligned}$$

□

Example 5.3. If μ is a self-decomposable distribution, then the Lévy measure admits always a Lebesgue density of the form $\nu(\mathrm{d}x) = \frac{k(x)}{|x|}\mathrm{d}x$ where the so-called k-function k is increasing on the negative half line and decreasing on the positive half line. Hence, $k_s\mathbb{1}_{(0,\delta]} \in BV(\mathbb{R})$ is equivalent to $\|k\|_\infty < \infty$. The condition $\int_0^\delta \log(y^{-1})(Dk_s)^+(\mathrm{d}y) < \infty$ is always satisfied for self-decomposable distributions owing to $(Dk_s)^+ = 0$. As indicated in Trabs (2014) it is more generally sufficient if the quotient $(k_s(x+)-\alpha)/x$ is bounded from above uniformly in $x \in (0,\varepsilon]$ for some $\varepsilon > 0$, meaning that the largest slope of k_s near zero is bounded, to obtain the sharp bound of the decay rate.

With this lemma at hand we can prove Propsition 5.1.

Proof of Proposition 5.1. Denoting $\widetilde{\rho} := Dk_s$, we define the monotone decreasing functions $k_s^+(x) := \int_x^\delta \widetilde{\rho}^+(\mathrm{d}y)$ and $k_s^-(x) := \int_x^\delta \widetilde{\rho}^-(\mathrm{d}y)$ for $x \in (0,\delta]$. To verfiy equivalence of (i) and (ii), we conclude from $k_s(\delta+) - k_s(x+) = k_s^+(x) - k_s^-(x)$ that

$$\begin{aligned}\|k_s \mathbb{1}_{(0,\delta]}\|_{BV} - \|\widetilde{\rho}^+|_{(0,\delta]}\|_{TV} &= \|\widetilde{\rho}^-|_{(0,\delta]}\|_{TV} \\ &= \sup_{x\in(0,\delta]} k_s^-(x) = \sup_{x\in(0,\delta]} \big\{k_s(x+) + k_s^+(x) - k_s(\delta+)\big\}.\end{aligned}$$

Hence, $\|k_s\|_{BV} < \infty$ if and only if $k_s(0+) < \infty$, owing to $0 \leqslant k_s^+(0+) = \|\widetilde{\rho}^+\|_{TV}$.
The inclusion (i)⇒(iii) immediately follows from the lower bound in Lemma 5.2.
To show (iii)⇒(i), we first note that if $\sigma^2 > 0$ then $|\varphi(u)| \lesssim e^{-cu^2}, u \in \mathbb{R}$, for some constant $c > 0$ which contradicts (iii). So let $\sigma^2 = 0$ and $u > 0$ without loss of generality. We deduce similarly to (5.2) and for any $\tau \in (0,\delta)$ and any $u > 1/\tau$

$$\begin{aligned}|\varphi(u)| &\leqslant \exp\Big(\int_{1/u}^{\tau}\int_y^{\tau} \frac{1-\cos(ux)}{x}\mathrm{d}x\widetilde{\rho}^-(\mathrm{d}y) + k_s(\tfrac{1}{u}+)\int_1^{\tau u}\frac{\cos(x)-1}{x}\mathrm{d}x\Big) \\ &\leqslant \exp\Big(-\int_1^{\tau u}\frac{1-\cos(x)}{x}\mathrm{d}x\Big(k_s(\tfrac{1}{u}+) - \int_{1/u}^{\tau}\widetilde{\rho}^-(\mathrm{d}y)\Big)\Big) \\ &= \exp\Big(-\int_1^{\tau u}\frac{1-\cos(x)}{x}\mathrm{d}x\big(k_s(\delta+) - k_s^+(\tfrac{1}{u}) + k_s^-(\tau)\big)\Big). \end{aligned} \tag{5.4}$$

Using

$$\int_1^y \frac{1-\cos(x)}{x}\mathrm{d}x = \log y + \int_y^\infty \frac{\cos x}{x}\mathrm{d}x + c_1 \quad \text{with} \quad c_1 := -\int_1^\infty \frac{\cos x}{x}\mathrm{d}x > 0$$

and $\lim_{y\to\infty}\int_y^\infty \frac{\cos x}{x}\mathrm{d}x = 0$, we find a function $f\colon [1,\infty) \to \mathbb{R}$ such that for some $T > 1$

$$\int_1^y \frac{1-\cos(x)}{x}\mathrm{d}x = f(y)\log y \quad \text{and} \quad f(y) \in (\tfrac{1}{2}, 2) \text{ for all } y \geqslant T.$$

Combining the lower bound on $|\varphi(u)|$ in Assumption (iii) and the upper bound (5.4), we conclude

$$\log c - \alpha\log 2 - \alpha\log u \leqslant \log|\varphi(u)| \leqslant -f(\tau u)\log(\tau u)\big(k_s(\delta+) - k_s^+(\tfrac{1}{u}) + k_s^-(\tau)\big)$$

and thus for $c_2 := \alpha\log 2 - \log c$

$$\begin{aligned}&\log u\big(f(\tau u)\big(k_s(\delta+) - k_s^+(\tfrac{1}{u}) + k_s^-(\tau)\big) - \alpha\big) \\ &\qquad \leqslant c_2 + \log(\tau^{-1})f(\tau u)\big(k_s(\delta+) - k_s^+(\tfrac{1}{u}) + k_s^-(\tau)\big).\end{aligned}$$

That implies for $u \geqslant T/\tau > 1$

$$\log u\big(\tfrac{1}{2}k_s(\delta+) - 2\|\widetilde{\rho}^+|_{(0,\delta]}\|_{TV} + \tfrac{1}{2}k_s^-(\tau)\big) - \alpha\big) \leqslant c_2 + 2\log(\tau^{-1})\big(k_s(\delta+) + k_s^-(\tau)\big).$$

Since the right-hand side is independent of u and $\log u \to \infty$ as $u \to \infty$, we obtain $\frac{1}{2}k_s(\delta+) - 2\|\widetilde{\rho}^+|_{(0,\delta]}\|_{TV} + \frac{1}{2}k_s^-(\tau) - \alpha < 0$ and therefore

$$\|\widetilde{\rho}^-|_{(0,\delta]}\|_{TV} = \sup_{\tau\in(0,\delta)} k_s^-(\tau) \leqslant 2\alpha + 4\|\widetilde{\rho}^+|_{(0,\delta]}\|_{TV} < \infty,$$

implying $\|k_s\mathbb{1}_{(0,\delta]}\|_{BV} < \infty$. □

Remark 5.4. With Assumption 5.A and Condition (iii) in Proposition 5.1, we obtain an upper bound for $\|k_s \mathbb{1}_{(0,\delta]}\|_{BV}$ which depends only on α and $\|(Dk_s)^+|_{(0,\delta]}\|_{TV}$.

Now we are in position to state the main result of this section.

Theorem 5.5. *Let μ be an infinitely divisible distribution with characteristic function φ satisfying Assumption 5.A. If and only if one (and thus all) of the conditions (i) to (iii) of Proposition 5.1 are satisfied, $1/\varphi$ is a Fourier multiplier on Besov spaces: There exists some $\alpha > 0$ such that for all $s \in \mathbb{R}, 1 \leqslant p, q \leqslant \infty$ the linear map*

$$B^{s+\alpha}_{p,q}(\mathbb{R}) \ni f \mapsto \mathcal{F}^{-1}\Big[\frac{\mathcal{F} f}{\varphi}\Big] \in B^s_{p,q}(\mathbb{R})$$

is bounded.

Proof. If $1/\varphi$ is a Fourier multiplier, then (iii) in Proposition 5.1 has to be fulfilled as carried out above.

Now, let the assumptions of Proposition 5.1 be satisfied. Using $\sigma^2 = 0$ and noting that $k_s \mathbb{1}_{(0,\delta]} \in BV(\mathbb{R})$ implies boundedness of k_s and thus $\int_0^1 x\nu_s(\mathrm{d}x) < \infty$, the characteristic function of μ can be represented by

$$\varphi(u) = \exp\Big(i\gamma_0 u + \int (e^{iux} - 1)\nu(\mathrm{d}x)\Big) = \varphi_c(u)\varphi_p(u), \quad \text{for} \quad u \in \mathbb{R},$$

where $\gamma_0 = \gamma - \int_0^1 x\nu_s(\mathrm{d}x) \in \mathbb{R}$ is a drift parameter and

$$\varphi_c(u) := \exp\Big(\int_{[-\delta,\delta]} (e^{iux} - 1)\nu(\mathrm{d}x)\Big), \quad \varphi_p(u) := \exp\Big(i\gamma_0 u + \int_{\mathbb{R}\setminus[-\delta,\delta]} (e^{iux} - 1)\nu(\mathrm{d}x)\Big).$$

Defining $\mu_c := \mathcal{F}^{-1}[\varphi_c]$ and $\mu_c := \mathcal{F}^{-1}[\varphi_p]$, this yields the decomposition $\mu = \mu_c * \mu_p$ into a convolution of an infinitely divisible distribution with compactly supported jump measure and distribution of compound Poisson type. The deconvolution operator decomposes into a composition $\mathcal{F}^{-1}[1/\varphi] = \mathcal{F}^{-1}[1/\varphi^c] * \mathcal{F}^{-1}[1/\varphi^p]$, where

$$\mathcal{F}^{-1}[1/\varphi_p] = \delta_{-\gamma_0} * \Big(e^{\nu(\mathbb{R}\setminus[-\delta,\delta])} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \big(\nu|_{(\mathbb{R}\setminus[-\delta,\delta])}\big)^{*k}\Big)$$

is a finite signed measure. Since convolution with a finite signed measure is a bounded map on $L^p(\mathbb{R})$ for any $0 < p \leqslant \infty$, we conclude from the Littlewood–Paley representation of Besov spaces that $B^s_{p,q}(\mathbb{R}) \ni f \mapsto \mathcal{F}^{-1}[1/\varphi_p] * f \in B^s_{p,q}(\mathbb{R})$ is a bounded linear map. For φ_c we use $\varphi_c'(u) = i\,\mathcal{F}[x\nu|_{[-\delta,\delta]}](u)\varphi_c(u)$ and $|\,\mathcal{F}[x\nu|_{[-\delta,\delta]}](u)| \lesssim (1+|u|)^{-1}$ by the bounded variation of $x\nu$ near the origin. The polynomial decay of φ_c implies then for some $\alpha > 0$

$$\begin{aligned}(1+|u|)^{-\alpha}|\varphi_c^{-1}(u)| &\lesssim 1 \quad \text{and}\\ (1+|u|)^{-\alpha}|(\varphi_c^{-1})'(u)| = (1+|u|)^{-\alpha}|\,\mathcal{F}[x\nu|_{[-\delta,\delta]}](u)/\varphi_c(u)| &\lesssim (1+|u|)^{-1}.\end{aligned}$$

Therefore, we can apply Corollary 4.11 by Girardi and Weis (2003) to conclude that $(1+|u|)^{-\alpha}/\varphi_c(u)$ is a Fourier muliplier on all Besov spaces $B^s_{p,q}(\mathbb{R})$ for all $s \in \mathbb{R}, 1 \leqslant p, q \leqslant \infty$. The assertion follows because $f \mapsto \mathcal{F}^{-1}[(1+iu)^{\alpha}\,\mathcal{F} f]$ is an isomorphism from $B^{s+\alpha}_{p,q}(\mathbb{R})$ onto $B^s_{p,q}(\mathbb{R})$. □

5.2. Discrete observations of the process

Let us observe $n \in \mathbb{N}$ increments of the Lévy process L at equidistant time points with observation distance $\Delta > 0$:

$$Y_k := L_{\Delta k} - L_{\Delta(k-1)}, \quad k = 1, \dots, n.$$

We will focus on low-frequency observations, meaning that Δ is fixed. Nevertheless, we track Δ in all estimates. The law of Y_k will be denoted by P_Δ. Using the empirical characteristic function $\varphi_{\Delta,n}(u) = \frac{1}{n}\sum_{k=1}^n e^{iuY_k}$, we obtain an empirical version $\widehat{\psi}_n''$ of ψ'' from (5.1) and thus we define

$$\widehat{\nu}_h(t) := -t^{-2}\,\mathcal{F}^{-1}\Big[\widehat{\psi}_n''(u)\,\mathcal{F}K(hu)\Big](t), \qquad t \neq 0,$$

where K is a band-limited kernel with bandwidth $h > 0$. Note that this estimator depends neither on the unknown volatility σ^2 nor on the drift parameter γ. The distribution function can then be estimated via the left and the right tail integrals

$$\widehat{N}_h(t) = -\int g_t(x)\,\mathcal{F}^{-1}\Big[\widehat{\psi}_n''(u)\,\mathcal{F}K(hu)\Big](x)\mathrm{d}x \quad \text{with} \quad g_t(x) := \begin{cases} x^{-2}\mathbb{1}_{(-\infty,t]}, & t < 0, \\ x^{-2}\mathbb{1}_{[t,\infty)}, & t > 0. \end{cases} \tag{5.5}$$

In contrast to the deconvolution estimator in (4.5), the estimator $\widehat{N}_h$ is always well defined owing to $g_t \in L^1(\mathbb{R})$. Given the distribution function estimator $\widehat{N}_h$, we construct the quantile estimator as the M-estimator

$$\widehat{q}^{\pm}_{\tau,h} := \operatorname{argmin}_{t\in[\eta_n,\infty)} |\widehat{N}_h(\pm t) - \tau|$$

for logarithmically decreasing threshold values $\eta_n \downarrow 0$. Compared to the deconvolution estimator, we minimize the contrast function over a different set. On the one hand we do not need a bounded interval and on the other hand it is bounded away from zero, due to the possible singularity at zero. Hence, we estimate the values $q^{\pm}_\tau \vee \eta_n$ instead of the generalized quantiles themselves. If $\tau \geqslant \nu(\mathbb{R}^{\pm})$, then $\widehat{q}^{\pm}_\tau \to 0$ in probability.

Throughout we suppose that the kernel function satisfies the following conditions for some order $p \in \mathbb{N}$

$$\begin{gathered} \int_{\mathbb{R}} K(x)\mathrm{d}x = 1, \qquad \int x^l K(x)\mathrm{d}x = 0 \quad \text{for } l = 1, \dots, p, \\ \operatorname{supp}\mathcal{F}K \subseteq [-1,1], \quad x^{p+1}K(x) \in L^1(\mathbb{R}). \end{gathered} \tag{5.6}$$

We distinguish between polynomially decaying characteristic functions and exponentially decaying characteristic functions. Define for some open set $U \subseteq \mathbb{R}$ and constants $\alpha, \delta, m, s, r, R > 0$ the classes of Lévy triplets

$$\begin{aligned} \mathcal{C}^s(m,U,R) := \Big\{(\sigma^2,\gamma,\nu)\Big|\sigma^2 \in [0,R], \gamma \in \mathbb{R}, \|x^m\nu\|_{L^1} \leqslant R, \\ \nu \text{ has a Lebesgue density on } U \text{ with } \|\nu\|_{C^s(U)} \leqslant R\Big\}, \end{aligned}$$

$$\mathcal{D}^s(\alpha, m, U, R) := \Big\{(0, \gamma, \nu) \in \mathcal{C}^s(m, U, R) \Big| \|(1 + |\bullet|)^{-\Delta\alpha}/\varphi_\Delta\|_\infty \leqslant R, \|x\nu\|_\infty \leqslant R,$$
$$x\nu \text{ satisfies Assumption 5.A with } \|(Dk_s)^+|_{(0,\delta]}\|_{TV} \leqslant R\Big\},$$
$$\mathcal{E}^s(\beta, m, U, r, R) := \Big\{(\sigma^2, \gamma, \nu) \in \mathcal{C}^s(m, U, R) \Big| \|\exp(-r\Delta|\bullet|^\beta)/\varphi_\Delta\|_\infty \leqslant R\Big\}. \tag{5.7}$$

As we will see the class $\mathcal{D}^s(\alpha, m, U, R)$ corresponds to mildly ill-posed estimation problems. Proposition 5.1 shows that the polynomial decay of the characteristic function together with the regularity Assumption 5.A implies already that $x\nu$ is bounded near zero. Hence, $\|x\nu\|_\infty \leqslant R$ is only an additional tail estimate. The estimation problem in the class $\mathcal{E}^s(\beta, m, U, R)$ is severely ill-posed leading to logarithmic convergence rates. The parameter β is closely related to the Blumenthal and Getoor (1961) index. Because the rates are so slow, we only need very mild assumptions in $\mathcal{E}^s(\beta, m, U, r, R)$.

One building block of the analysis of the quantile estimator in the deconvolution model in Chapter 4 was a convergence result for the risk of the density estimator under uniform loss. When we determine convergence rates for $\widehat{\nu}_h$, we thus consider again uniform loss.

Proposition 5.6. *Let $\alpha, \beta, s, r, R > 0, m > 4$ and let the kernel satisfy (5.6) with order $p \geqslant s$ and let $U \subseteq \mathbb{R}$ be a bounded, open set which is bounded away from zero. Then we have*

(i) uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{D}^s(\alpha, m, U, R)$ for $h = h_{n,\Delta} = (\frac{\log n\Delta}{n\Delta})^{1/(2s+2\Delta\alpha+1)}$

$$\sup_{t \in U} |\widehat{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{D}^s}\Big(\Big(\frac{\log n\Delta}{n\Delta}\Big)^{s/(2s+2\Delta\alpha+1)}\Big),$$

(ii) uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{E}^s(\beta, m, U, r, R)$ for $h = h_{n,\Delta} = (\frac{\delta \log n\Delta}{2r\Delta})^{-1/\beta}$ with $\delta \in (0, 3/2)$

$$\sup_{t \in U} |\widehat{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{E}^s}\Big(\Big(\frac{\log n\Delta}{\Delta}\Big)^{-s/\beta}\Big).$$

In the mildly ill-posed case the rates correspond to the deconvolution problem with an error distribution whose characteristic function decays with polynomial rate $\Delta\alpha$, see Proposition 4.6. It can easily be verified that the for the pointwise loss we have the same rates without the logarithm in (i). They coincide with the convergence rates for the pointwise loss by (Kappus, 2014, Thm. 3.5), who has considered only finite variation Lévy processes. Kappus (2012) shows that these are minimax optimal. In the classical density estimation the logarithm is known to be unavoidable for the uniform loss.

The distribution function estimator $\widehat{N}_h$ was studied by Nickl et al. (2013) in a high-frequency regime. For low frequency observations a modification of $\widehat{N}_h$ was considered by Nickl and Reiß (2012). In both articles a uniform central limit theorem has been established. To this end, assumptions have been imposed which ensure that the parametric rate can be attained. Therefore, it is of interest to derive general convergence rates for $\widehat{N}_h$.

Proposition 5.7. *Let $U \subseteq \mathbb{R}$ be an open set, $t \in U$ and $\alpha, \beta, s, r, R > 0$ and $m > 4$. Suppose the kernel satisfies (5.6) with order $p \geqslant s + 1$. Let $\Delta \in (0, 1)$ and $n \to \infty$. Then:*

(i) *The choice* $h = h_{n,\Delta} = (n\Delta)^{-1/(2s+(2\Delta\alpha\vee 1)+1)}$ *yields uniformly in* $(\sigma^2,\gamma,\nu) \in \mathcal{D}^s(\alpha,m,U,R)$

$$|\widehat{N}_h(t)-N(t)| = \mathcal{O}_{P,\mathcal{D}^s}(r_{n,\Delta}), \quad r_{n,\Delta} := \begin{cases} (n\Delta)^{-(s+1)/(2s+2\Delta\alpha+1)}, & \text{for } \Delta\alpha > 1/2, \\ (n\Delta)^{-1/2}(\log n\Delta)^{1/2}, & \text{for } \Delta\alpha = 1/2, \\ (n\Delta)^{-1/2}, & \text{for } \Delta\alpha \in (0,1/2). \end{cases}$$

(ii) *The choice* $h = h_{n,\Delta} = (\frac{\delta}{2r})^{-1/\beta}(\frac{\log(n\Delta)}{\Delta})^{-1/\beta}$, *for any* $\delta \in (0,3/2)$, *yields uniformly in* $(\sigma^2,\gamma,\nu) \in \mathcal{E}^s(\beta,m,U,r,R)$

$$|\widehat{N}_h(t)-N(t)| = \mathcal{O}_{P,\mathcal{E}^s}\Big(\Big(\frac{\log(n\Delta)}{\Delta}\Big)^{-(s+1)/\beta}\Big).$$

Remark 5.8. This result could be strengthened to the uniform loss on $\mathbb{R}\setminus[-\eta,\eta]$ for any $\eta > 0$, cf. (5.42) below.

As expected, the convergence rates for $\widehat{N}_n$ are faster than for density estimation because we gain one degree of smoothness. In particular, we achieve the parametric rate for a Lévy process with very slowly decaying characteristic function or for Δ sufficiently small.

To analyze the estimation error of the quantile estimators, we use a Taylor expansion similarly to the deconvolution case considered in Chapter 4. Due to the continuity of the inverse Fourier transform, the derivative $\widehat{N}_h'$ is well defined as $\widehat{N}_h'(t) = -\operatorname{sign}(t)\widehat{\nu}_h(t)$ for $t \neq 0$. Hence,

$$0 \approx \widehat{N}_h(\widehat{q}^+_{\tau,h}) - \tau = \widehat{N}_h(q^+_\tau) - N(q^+_\tau) - (\widehat{q}^+_{\tau,h} - q^+_\tau)\widehat{\nu}_h(\xi^+)$$

for some intermediate point ξ^+ between $q^+_\tau$and $\widehat{q}^+_{\tau,h}$ and similarly for $\widehat{q}^-_{\tau,h}$. By continuity of $\widehat{N}_h$ the probability of the event $\{\widehat{N}_h(\widehat{q}^+_{\tau,h}) - \tau = 0\}$ converges to one, similarly to the deconvolution case. On this event the estimation error can therefore be represented as

$$|\widehat{q}^\pm_{\tau,h} - q^\pm_\tau| = \Big|\frac{\widehat{N}_h(\pm q^\pm_\tau) - N(\pm q^\pm_\tau) + \tau - \widehat{N}_h(\widehat{q}^+_{\tau,h})}{\widehat{\nu}_h(\xi^\pm)}\Big| = \frac{|\widehat{N}_h(\pm q^\pm_\tau) - N(\pm q^\pm_\tau)|}{|\widehat{\nu}_h(\xi^\pm)|}, \quad (5.8)$$

for intermediate points $\xi^\pm$. Since the convergence rates for the the density estimator $\widehat{\nu}_h$ and for the distribution function estimator $\widehat{N}_h$ are already established, it remains to show consistency the quantile estimator $\widehat{q}^\pm_{\tau,h}$ itself. To this end, a minimal global regularity of ν is required (in Chapter 4 boundedness of the density was imposed). We need to specify our nonparametric classes further such that the quantiles exist. Writing

$C^{s'}(\mathbb{R}) = B^{s'}_{\infty,\infty}(\mathbb{R})$ for $s' \in (-1, 0]$ and with $\zeta, \eta > 0$, we define for a given $\tau > 0$

$$\begin{aligned}&\widetilde{\mathcal{D}}^{s,s'}_\tau(\alpha, m, \zeta, \eta, R) \qquad (5.9)\\ &:= \Big\{(\sigma^2,\gamma,\nu)\Big|\exists q_\tau^+, q_\tau^- \in (\eta,\infty) : \int_{-\infty}^{-q_\tau^-} \mathrm{d}\nu = \tau = \int_{q_\tau^+}^{\infty} \mathrm{d}\nu, \|\nu\|_{C^{s'}(\mathbb{R}\setminus[-\eta,\eta])} < R,\\ &\qquad (\sigma^2,\gamma,\nu) \in \mathcal{D}^s(\alpha, m, (q_\tau^+ - \zeta, q_\tau^+ + \zeta) \cup (q_\tau^- - \zeta, q_\tau^- + \zeta), R), \nu(q_\tau^\pm) > \tfrac{1}{R}\Big\},\\ &\widetilde{\mathcal{E}}^{s,s'}_\tau(\beta, m, \zeta, \eta, r, R) \qquad (5.10)\\ &:= \Big\{(\sigma^2,\gamma,\nu)\Big|\exists q_\tau^+, q_\tau^- \in (\eta,\infty) : \int_{-\infty}^{-q_\tau^-} \mathrm{d}\nu = \tau = \int_{q_\tau^+}^{\infty} \mathrm{d}\nu, \|\nu\|_{C^{s'}(\mathbb{R}\setminus[-\eta,\eta])} < R,\\ &\qquad (\sigma^2,\gamma,\nu) \in \mathcal{E}^s(\beta, m, (q_\tau^+ - \zeta, q_\tau^+ + \zeta) \cup (q_\tau^- - \zeta, q_\tau^- + \zeta), r, R), \nu(q_\tau^\pm) > \tfrac{1}{R}\Big\}.\end{aligned}$$

As in the deconvolution setting we obtain the same rates for quantile estimation as for distribution function estimation.

Theorem 5.9. *Let $\tau > 0$ and $\alpha, \beta, s, \zeta, r, R > 0, s' \in (-1, 0]$ and $m > 4$. Suppose the kernel satisfies (5.6) with order $p \geqslant s + 1$ and $\eta_n \downarrow 0$ with $\eta_n^{-1} \lesssim \log n$. Then we obtain for $\Delta \in (0, 1)$ and $n \to \infty$:*

(i) The choice $h = h_{n,\Delta} = (n\Delta)^{-1/(2s+(2\Delta\alpha\vee 1)+1)}$ yields uniformly in $(\sigma^2, \gamma, \nu) \in \widetilde{\mathcal{D}}^{s,s'}_\tau(\alpha, m, \zeta, \eta_n, R)$

$$|\widehat{q}^{\pm}_{\tau,h} - q^\pm_\tau| = \mathcal{O}_{P,\widetilde{\mathcal{D}}^{s,s'}_\tau}(r_{n,\Delta}), \quad r_{n,\Delta} := \begin{cases} (n\Delta)^{-(s+1)/(2s+2\Delta\alpha+1)} & \text{for } \Delta\alpha > 1/2,\\ (n\Delta)^{-1/2}(\log n\Delta)^{1/2} & \text{for } \Delta\alpha = 1/2,\\ (n\Delta)^{-1/2} & \text{for } \Delta\alpha \in (0, 1/2).\end{cases}$$

(ii) The choice $h = h_{n,\Delta} = (\frac{\delta}{2r})^{-1/\beta}(\frac{\log(n\Delta)}{\Delta})^{-1/\beta}$, for any $\delta \in (0, 2/3)$, yields uniformly in $(\sigma^2, \gamma, \nu) \in \widetilde{\mathcal{E}}^{s,s'}_\tau(\beta, m, \zeta, \eta_n, r, R)$

$$|\widehat{q}^{\pm}_{\tau,h} - q^\pm_\tau| = \mathcal{O}_{P,\widetilde{\mathcal{E}}^{s,s'}_\tau}\Big(\Big(\frac{\log(n\Delta)}{\Delta}\Big)^{-(s+1)/\beta}\Big).$$

Remark 5.10. For high-frequency data, that is for $\Delta \to 0$, we obtain in the class $\widetilde{\mathcal{D}}^{s,s'}_\tau(\alpha, \zeta, \eta_n, R)$ always the parametric rate $(n\Delta)^{-1/2}$. In $\widetilde{\mathcal{E}}^{s,s'}_\tau(\beta, \zeta, \eta_n, r, R)$ one should choose instead $h_{n,\Delta} = \Delta^{1/\beta}$ such that our estimates in the proof of Theorem 5.9 yield the almost optimal rate $(n\Delta)^{-1/2}|\log \Delta|$ provided that s is large enough such that the bias condition $\Delta^{(s+1)/\beta} = \mathcal{O}((n\Delta)^{-1/2})$ is satisfied and for $\beta \in (4/3, 2]$, thus especially for $\sigma^2 > 0$. With stronger assumptions the latter restriction can be circumvented and the rate can be improved. In fact, in Nickl et al. (2013) we show that the parametric rate $(n\Delta)^{-1/2}$ can be obtained with this estimator under suitable conditions on ν.

5.3. Financial example revisited: Observation of option prices

5.3.1. Quantile estimator and convergence rates

Let us consider again the exponential Lévy model for asset prices from Chapter 2

$$S_t = S_0 e^{rt+L_t}, \quad t \geqslant 0, \qquad (5.11)$$

with initial value $S_0 > 0$, riskless interest rate $r \geqslant 0$ and with the driving Lévy process L whose characteristic triplet is (σ^2, γ, ν). Throughout this section we consider only Lévy processes which satisfy the martingale condition, cf. (2.4),

$$\frac{\sigma^2}{2} + \gamma + \int_{-\infty}^{\infty} (e^x - 1 - x\mathbb{1}_[- 1, 1](x))\nu(\mathrm{d}x) = 0.$$

Above we have discussed the calibration of the model. Another important objective for pricing and hedging is to estimate the risk within the model. There are many possibilities to measure the risk. One of the most popular quantities is the value-at-risk at some level $\tau \in (0,1)$ which is given by the $(1-\tau)$-quantile of the distribution of the loss of the asset. The generalized quantiles of ν under the risk neutral measure are a closely related concept which takes only the influence of shocks, in the sense of large jumps, into account. They may also be useful for dynamic quantile hedging in the spirit of Föllmer and Leukert (1999).

Recall the definition of the option function O from (2.5) based on vanilla options with some fixed maturity T. We observe O at a finite number of (transformed) strike prices $x_1, \dots, x_n$, for $n \in \mathbb{N}$, corrupted by noise

$$O_j = O(x_j) + \delta_j \varepsilon_j \quad j = 1, \dots, n, \tag{5.12}$$

where (ε_j) are i.i.d. centered random variables with $\operatorname{Var}(\varepsilon_j) = 1$ and with local noise levels (δ_j). By interpolating the observations $(x_j, O_j)_{j=1,\dots,n}$, we construct an empirical version $\widetilde{O}$ of the option function as in Section 2.2. Since the function O is related to the characteristic function φ_T of X_T via the pricing formula (2.7), we obtain an estimator of φ_T given by

$$\widetilde{\varphi}_{T,n}(u) := 1 - u(u+i)\,\mathcal{F}\,\widetilde{O}(u+i). \tag{5.13}$$

Using $\widetilde{\varphi}_{T,n}$, we obtain a quantile estimator as described at the beginning of this chapter:

(i) The second derivative of the characteristic exponent can be estimated by differentiating (5.13) twice. We define

$$\begin{aligned}\widetilde{\psi}_n'(u) &:= -\frac{(2u+i)\,\mathcal{F}\,\widetilde{O}(u+i) + u(iu-1)\,\mathcal{F}[x\widetilde{O}](u+i)}{T(1-u(u-i)\,\mathcal{F}\,\widetilde{O}(u+i))},\\ \widetilde{\psi}_n''(u) &:= -\frac{(2\,\mathcal{F}\,\widetilde{O}(u+i) + (4iu-2)\,\mathcal{F}[x\widetilde{O}](u+i) - (u^2+iu)\,\mathcal{F}[x^2\widetilde{O}](u+i))}{T(1-u(u-i)\,\mathcal{F}\,\widetilde{O}(u+i))}\\ &\qquad - \frac{1}{T}\Big(\frac{(2u+i)\,\mathcal{F}\,\widetilde{O}(u+i) + u(iu-1)\,\mathcal{F}[x\widetilde{O}](u+i)}{1-u(u-i)\,\mathcal{F}\,\widetilde{O}(u+i)}\Big)^2.\end{aligned} \tag{5.14}$$

Using a kernel K with bandwidth $h > 0$ satisfying (5.6), we obtain the density estimator

$$\widetilde{\nu}_h(t) := -t^{-2}\,\mathcal{F}^{-1}\Big[\widetilde{\psi}_n''(u)\,\mathcal{F}\,K(hu)\Big](t), \qquad t \neq 0.$$

(ii) Integrating $\widetilde{\nu}_h$, the estimator of the generalized distribution function is given by

$$\widetilde{N}_h(t) = -\int g_t(x)\,\mathcal{F}^{-1}\Big[\widetilde{\psi}_n''(u)\,\mathcal{F}\,K(hu)\Big](x)\mathrm{d}x, \quad t \neq 0,$$

with g_t from (5.5).

(iii) For $\tau > 0$ the quantile estimators are defined as the minimum contrast estimators

$$\widetilde{q}^{\pm}_{\tau,h} := \operatorname{argmin}_{t\in[\eta_n,\infty)} |\widetilde{N}_h(\pm t) - \tau|$$

with threshold value $\eta_n \downarrow 0$.

These estimators are well defined on the event $A = \{\forall u \in [-1/h, 1/h] : \widetilde{\varphi}_{T,n}(u) \neq 0\}$ whose probability increases if $\widetilde{\varphi}_{T,n}$ concentrates around the true φ_T. In an idealized model Söhl (2010) shows $P(A) = 1$. Note that $\widetilde{\nu}_h$ is different from the finite activity estimator in Chapter 2 which have been proposed by Belomestny and Reiß (2006a). The k-function estimator from Trabs (2014) in the self-decomposable model relies on a similar idea but using only the first derivative of ψ.

Since we want to concentrate on the main aspects in the error analysis and to avoid technicalities, we will work in the idealized Gaussian white noise model which was considered by Söhl (2014) as well. Assume that the noise levels of the observations (5.12) are given by the values $\delta_j = \delta(x_j)$, $j = 1,\dots,n$, of some function $\delta : \mathbb{R} \to \mathbb{R}_+$. The observed strike prices are assumed to be the quantiles $x_j = F^{-1}(j/(n+1))$, $j = 1,\dots,n$, of a distribution with distribution function $F : \mathbb{R} \to [0,1]$ and density $f > 0$. Incorporating the observation errors as well as their distribution, we define the general noise level

$$\varrho(x) = \delta(x)/\sqrt{f(x)}. \tag{5.15}$$

Instead of assuming that the observation points are given by the quantiles of f one may also assume that the observation points are sampled randomly from the density f. For standard normally distributed (ε_j) Brown and Low (1996) have shown asymptotic equivalence in the sense of Le Cam of the nonparametric regression model (5.12) and the Gaussian white noise model

$$\mathrm{d}Z(x) = O(x)\mathrm{d}x + n^{-1/2}\varrho(x)\mathrm{d}W(x)$$

with a two-sided Brownian motion W for x on a possibly growing bounded interval. Grama and Nussbaum (2002) have extended the equivalence to more general error distributions. More details on this equivalence can be found in the papers by Söhl (2014) and by Trabs (2014, Supplement).

Z is an empirical version of the antiderivative of O. In that sense we define $\mathcal{F}\widetilde{O}(u) := \mathcal{F}[\mathrm{d}Z](u) = \mathcal{F}O(u) + n^{-1/2}\int e^{iux}\varrho(x)\mathrm{d}W(x)$ and analogously for $\mathcal{F}[x\widetilde{O}]$ and $\mathcal{F}[x^2\widetilde{O}]$. Owing to (5.13), the estimation error of the $\widetilde{\varphi}_{T,n}$ is given by the Gaussian process

$$\Phi_n(u) := (\widetilde{\varphi}_{T,n} - \varphi_T)(u) = u(u+i)\,\mathcal{F}[O-\widetilde{O}](u+i) = n^{-1/2}u(u-i)\int e^{iux-x}\rho(x)\mathrm{d}W(x).$$

Applying Dudley's theorem, we obtain the following path property of Φ_n. This lemma is in line with the results by Söhl (2010) and Proposition 1 in Söhl (2014).

Lemma 5.11. *Grant $\int(1+|x|)^m e^{-2x}\rho^2(x)\mathrm{d}x < \infty$ for some $m > 4$. Then Φ_n is twice $L^2(P)$-differentiable with derivatives*

$$\Phi_n^{(1)}(u) := n^{-1/2}\int e^{iux-x}(2u - i + (iu^2 + u)x)\rho(x)\mathrm{d}W(x), \tag{5.16}$$

$$\Phi_n^{(2)}(u) := n^{-1/2}\int e^{iux-x}(2 + 2(2iu+1)x + u(i-u)x^2)\rho(x)\mathrm{d}W(x).$$

Moreover, $\Phi_n^{(0)} := \Phi_n, \Phi_n^{(1)}, \Phi_n^{(2)}$ *have versions that are almost surely continuous and satisfy for any* $U > 0$

$$\mathbb{E}\left[\|\Phi_n^{(k)}\|_{L^\infty[-U,U]}\right] \lesssim n^{-1/2}U^2\sqrt{\log U} \quad \text{for} \quad k = 0,1,2.$$

In the following we may use these almost surely continuous and bounded versions. Let us first state a result on the uniform loss for the density estimator $\widetilde{\nu}_h$.

Proposition 5.12. *Let* $\alpha, \beta, s, r, R > 0$ *and let the kernel satisfy (5.6) with order* $p \geqslant s$ *and let* $U \subseteq \mathbb{R}$ *be a bounded, open set which is bounded away from zero. Suppose* $\|(1 \vee x^2)e^{-x}\rho(x)\|_\infty < \infty$. *Then we have*

(i) uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{D}^s(\alpha, 2, U, R)$ *with* $h = h_n = \left(\frac{\log n}{n}\right)^{1/(2s+2T\alpha+5)}$

$$\sup_{t \in U} |\widetilde{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{D}^s}\Big(\Big(\frac{\log n}{n}\Big)^{s/(2s+2T\alpha+5)}\Big),$$

(ii) uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{E}^s(\beta, 2, U, r, R)$ *with* $h = h_n = \left(\frac{1}{4r}\log n\right)^{-1/\beta}$

$$\sup_{t \in U} |\widetilde{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{E}^s}\big((\log n)^{-s/\beta}\big).$$

The convergence rates for the pointwise loss are the same without the logarithmic factor in (i). They coincide with the rates by Belomestny and Reiß (2006a) who have considered only the extreme cases: If $\sigma^2 = 0$ and ν is a finite measure, the pointwise risk converges with rate $n^{-s(2s+5)}$, and if $\sigma^2 > 0$, we obtain the rate $(\log n)^{-s/2}$. The estimator for the k-function by Trabs (2014) achieves the same rate as the corresponding pointwise result in Proposition 5.12(i). Since in the two afore mentioned papers lower bounds have been proved and the logarithm is unavoidable for uniform loss, the above rates appear to be minimax optimal.

Recalling the function classes from (5.9), we obtain the following convergence rates for the quantile estimators $\widetilde{q}_{\tau,h}^{\pm}$.

Theorem 5.13. *Let* $\tau > 0$ *and* $\alpha, \beta, s, \zeta, r, R > 0, s' \in (-1, 0]$. *Suppose that* $\|(1 \vee x^2)e^{-x}\rho\|_\infty \lesssim 1$ *and* $\int (1+|x|)^m e^{-2x}\rho^2(x)\mathrm{d}x < \infty$ *for some* $m > 4$. *Suppose the kernel satisfies (5.6) with order* $p \geqslant s+1$ *and let* $\eta_n \downarrow 0$ *with* $\eta_n^{-1} \lesssim \log n$. *Then we obtain for* $n \to \infty$*:*

(i) uniformly in $\widetilde{\mathcal{D}}_\tau^{s,s'}(\alpha, 2, \zeta, \eta_n, R)$ *with* $h = h_n = n^{-1/(2s+2T\alpha+5)}$

$$|\widetilde{q}_{\tau,h}^{\pm} - q_\tau^{\pm}| = \mathcal{O}_{P,\widetilde{\mathcal{D}}_\tau^{s,s'}}\big(n^{-(s+1)/(2s+2T\alpha+5)}\big),$$

(ii) uniformly in $\widetilde{\mathcal{E}}_\tau^{s,s'}(\beta, 2, \zeta, \eta_n, r, R)$ *with* $h = h_n = \left(\frac{1}{4r}\log n\right)^{-1/\beta}$

$$|\widetilde{q}_{\tau,h}^{\pm} - q_\tau^{\pm}| = \mathcal{O}_{P,\widetilde{\mathcal{E}}_\tau^{s,s'}}\big((\log n)^{-(s+1)/\beta}\big).$$

Compared to Theorem 5.9, the rates in (i) are always slower. In particular, the parametric rate can never be achieved. Heuristically, this is because we estimate a derived parameter of the state price density, which is basically the second derivative of the observed option function O. In the Fourier domain we see in the pricing formula (2.7) that $\mathcal{F}O$ decays two polynomial degrees faster than φ_T such that the ill-posedness of the statistical problem is larger. In the severely ill-posed case (ii) the rate is the same in both observation schemes since the rates are only logarithmically slow. The moment assumption is weakened to second moments, which are necessary for the identification identity (5.1). Instead the existence of fourth moments is implicitly imposed on the error distribution in the regression scheme. Although the theorem is stated for $\widetilde{\mathcal{D}}_\tau^{s,s'}(\alpha,2,\zeta,\eta,R)$, the boundedness of $x\nu$ is not needed here and could be dropped.

5.3.2. Data-driven choice of the bandwidth

Of course the optimal bandwidth is not known to the practitioner. To provide an adaptive method, we use again a Lepski-like method by considering a family of quantile estimators $\{\widetilde{q}_{\tau,h} : h \in \mathcal{B}_n\}$ for an appropriate set of bandwidths $\mathcal{B}_n$. Following the construction in Section 4.2, we define for a constant $L > 1$ and a sequence $(N_n) \subseteq \mathbb{N}$ satisfying $n^{-1}L^{N_n} \sim (\log n)^{-5}$

$$h_{n,j} := n^{-1}L^j \quad \text{for} \quad j = 0,\dots,N_n.$$

To ensure that the density estimator $\widetilde{\nu}_h$ is consistent for any $h \in \mathcal{B}_n$, we choose the minimal bandwidth via

$$\begin{aligned}\widetilde{j}_n := \min\Big\{& j = 0,\dots,N_n :\\ &\frac{1}{2} \leqslant \frac{(\log n)^2}{n^{1/2}}\Big(\int_{-1/h_{n,j}}^{1/h_{n,j}} \frac{(1+u^4)\mathbb{1}_{\{|\widetilde{\varphi}_{T,n}(u)|\geqslant(1+|u|)^2/n^{1/2}\}}}{|\widetilde{\varphi}_{T,n}(u)|^2}\mathrm{d}u\Big)^{1/2} \leqslant 1\Big\}\end{aligned}$$

and define

$$\mathcal{B}_n := \{h_{n,\widetilde{j}_n},\dots,h_{n,N_n}\}.$$

To choose the bandwidth from $\mathcal{B}_n$ which mimics the oracle bandwidth in Theorem 5.13, we have to estimate the standard deviation of the stochastic error. This problem is similar to the one considered by Söhl (2014) who determined the asymptotic distribution of the finite activity estimators from Chapter 2 and derived confidence sets. The stochastic error is dominated by its linearization and thus we define, cf. Lemma 5.26,

$$\begin{aligned}\widetilde{\Sigma}_{n,h}^{\pm} :=& \frac{1}{2\pi n^{1/2}T}\Big(\|x^2e^{-x}\rho(x)\|_\infty\|\widetilde{\chi}^{(2)}_{\widetilde{q}^{\pm}_{\tau,h}}\|_{L^2}\\ &+\|xe^{-x}\rho(x)\|_\infty\|\widetilde{\chi}^{(1)}_{\widetilde{q}^{\pm}_{\tau,h}}\|_{L^2}+\|e^{-x}\rho(x)\|_\infty\|\widetilde{\chi}^{(0)}_{\widetilde{q}^{\pm}_{\tau,h}}\|_{L^2}\Big)\end{aligned} \tag{5.17}$$

with auxiliary functions, for $u\in\mathbb{R}, t\neq 0$,

$$\begin{aligned}\widetilde{\chi}_t^{(0)}(u) &:= \mathcal{F}\, g_t(-u)\,\mathcal{F}\, K(hu)\Big(u(u-i)\frac{T^2\widetilde{\psi}_n'(u)^2 - T\widetilde{\psi}_n''(u)}{\widetilde{\varphi}_{T,n}(u)} \\ &\qquad + 2T(i-2u)\frac{\widetilde{\psi}_n'(u)}{\widetilde{\varphi}_{T,n}(u)} + 2\widetilde{\varphi}_{T,n}^{-1}(u)\Big),\\ \widetilde{\chi}_t^{(1)}(u) &:= \mathcal{F}\, g_t(-u)\,\mathcal{F}\, K(hu)\Big((4iu+2)\widetilde{\varphi}_{T,n}^{-1}(u) - 2Tu(iu+1)\frac{\widetilde{\psi}_n'(u)}{\widetilde{\varphi}_{T,n}(u)}\Big),\\ \widetilde{\chi}_t^{(2)}(u) &:= u(i-u)\,\mathcal{F}\, g_t(-u)\,\mathcal{F}\, K(hu)\widetilde{\varphi}_{T,n}^{-1}(u).\end{aligned}$$

Note that $\widetilde{\Sigma}_{n,h}^{\pm}$ are monotone decreasing in h. The magnitude of the stochastic error of $\widetilde{q}_{\tau,h}^{\pm}$ can then be estimated by

$$\widetilde{V}_n^{\pm}(h) := \frac{(1+\delta)\sqrt{2\log\log n}\widetilde{\Sigma}_{n,h}^{\pm}}{|\widetilde{\nu}_h(\widetilde{q}_{\tau,h}^{\pm})|} \tag{5.18}$$

for any small $\delta>0$. Defining

$$\mathcal{U}_h^{\pm} := [\widetilde{q}_{\tau,h}^{\pm} - \widetilde{V}_n^{\pm}(h), \widetilde{q}_{\tau,h}^{\pm} + \widetilde{V}_n^{\pm}(h)],$$

the adaptive estimator is defined as

$$\widetilde{q}_\tau^{\pm} := \widetilde{q}_{\tau,\widetilde{h}^{\pm}}^{\pm} \quad \text{with} \quad \widetilde{h}^{\pm} := \max\{h\in\mathcal{B}_n : \bigcap_{\mu\leqslant h,\mu\in\mathcal{B}_n}\mathcal{U}_h^{\pm}\neq\emptyset\}.$$

Theorem 5.14. *Let $\tau>0$ and $\alpha,\beta,s,\zeta,r,R>0, s'\in(-1,0]$. Suppose that $\|(1\vee x^2)e^{-x}\rho\|_\infty\lesssim 1$ and $\int(1+|x|)^m e^{-2x}\rho^2(x)\mathrm{d}x<\infty$ for some $m>4$. Suppose the kernel satisfies (5.6) with order $p\geqslant s+1$ and let $\eta_n\downarrow 0$ with $\eta_n^{-1}\lesssim\log n$. Then we obtain for $n\to\infty$:*

(i) uniformly in $\widetilde{\mathcal{D}}_\tau^{s,s'}(\alpha,2,\zeta,\eta_n,R)$

$$|\widetilde{q}_\tau^{\pm} - q_\tau^{\pm}| = \mathcal{O}_{P,\widetilde{\mathcal{D}}_\tau^{s,s'}}\big(((\log\log n)n^{-1})^{(s+1)/(2s+2T\alpha+5)}\big),$$

(ii) uniformly in $\widetilde{\mathcal{E}}_\tau^{s,s'}(\beta,2,\zeta,\eta_n,r,R)$

$$|\widetilde{q}_\tau^{\pm} - q_\tau^{\pm}| = \mathcal{O}_{P,\widetilde{\mathcal{E}}_\tau^{s,s'}}\big((\log n)^{-(s+1)/\beta}\big).$$

In the mildly ill-posed case (i) the adaptive method looses a $\log\log n$-factor compared to the oracle choice in Theorem 5.13. As we have argued in Section 4.2 this is unavoidable. In the severely ill-posed case (ii) the rates are already logarithmically slow such that the adaptive method causes no additional loss.

	10^2·RMSE		$\widetilde{q}_\tau^-$		$\widetilde{q}_\tau^-$	
τ	q_τ^-	q_τ^+	oracle	Lepski	oracle	Lepski
0.5	0.1778	0.1241	0.346	4.806	0.444	2.246
1.0	0.1201	0.0868	0.297	1.396	0.361	0.741
1.5	0.0929	0.0665	0.185	0.890	0.434	0.869
2.0	0.0726	0.0563	0.275	0.867	0.314	0.670
2.5	0.0624	0.0461	0.233	0.652	0.424	0.694

Table 5.1.: Empirical RMSE of the quantile estimators $\widetilde{q}_\tau^\pm$ from 1000 Monte Carlo simulations of the CGMY model.

5.3.3. Simulations and real data

We will illustrate the quantile estimation method in simulations from the CMGY model introduced by Carr et al. (2002). It is a generalization of the variance gamma model which we have studied in Chapter 2. The driving Lévy process of the asset is tempered stable and may have a diffusion component. For parameters $C > 0, M, G \geqslant 0$ and $Y < 2$ the Lévy measure in the CMGY model is given by the Lebesgue density

$$\nu_{cgmy}(x) = \begin{cases} C|x|^{-1-Y}e^{-G|x|}, & x < 0, \\ Cx^{-1-Y}e^{-Mx}, & x > 0. \end{cases}$$

With a fifth parameter $\sigma \geqslant 0$ the characteristic triplet of the underlying Lévy process is $(\sigma^2, \gamma, \nu_{cgmy})$ where the drift is determined by the martingale condition. For the simulations we set $C = 1, G = 5, M = 8, Y = 0.5$ and $\sigma = 0.1$ in view of the empirical results by Carr et al. (2002). The riskless interest rate is chosen as $r = 0.06$. As in Section 2.3 the design points $x_1, \ldots, x_n$ are constructed as $j/(n+1)$-quantiles of a $\mathcal{N}(0, 1/2)$ distribution. We simulate $n = 100$ option prices with time to maturity $T = 0.25$ that are three months. The local noise levels δ_j are chosen as 1% of the observed prices $O(x_j)$. The same 1% is assumed as a rule of thumb for the real data, where the noise level is unknown (cf. Cont and Tankov, 2004a, p. 439).

In order to apply the estimation procedure, we have to choose some parameters. The truncation value is set to $\eta = 0.02$. To construct the bandwidth set $\mathcal{B}_n$, we take $L = 1.1$. To compute $\widetilde{\Sigma}_{n,h}^\pm$, we need the noise function ϱ from (5.15). The density f of the distribution of the strikes is necessary, but in general not known to the practitioner. It can be estimated from the observation points $(x_j)_{j=1,\ldots,n}$ using some standard density estimation method. As in Söhl and Trabs (2014) we will apply a triangular kernel estimator, where the bandwidth is chosen by Silverman's rule of thumb.

To assess the performance of the estimation procedure, we compare the Lepski choice of the bandwidth to the oracle bandwidth, meaning that h is chosen such that the empirical root mean squared error (RMSE) is minimized. Our simulation results are summarized in Tabel 5.1 for $\tau \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$. Although the RMSE of the Lepski method is significantly larger than the oracle choice, the method achieves reasonable estimation errors. Note that the sample sizes are relatively small, cf. Table 2.1. We see that for smaller τ the estimation errors are larger. This is because small values of τ correspond to rare large jumps such that the jump density is small. Consequently, the

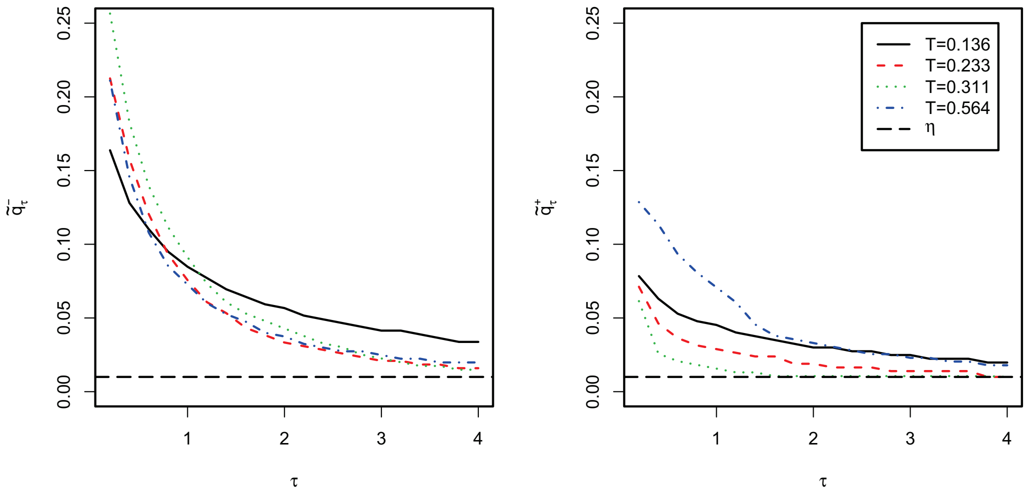

Figure 5.1.: Estimated generalized quantiles of negative jumps *(left)* and positive jumps *(right)* based on option prices from May 29, 2008, with four maturities T.

estimation error (5.8) is large. Since the stochastic estimation error has to be estimated by $\widetilde{V}_n^{\pm}$ from (5.18), this effect is more severe for the Lepski method.

Let us finally apply the estimation method to prices of DAX options from May 29, 2008. This data set[1] has already been studied in Chapter 2. Figure 5.1 shows the estimated quantiles $\widetilde{q}_\tau^{\pm}$ for four different maturities between two and seven months and for $\tau \in \{0.2, 0.4, \dots, 4\}$. Due to this finer grid, the threshold value is set to $\eta = 0.01$. Although not ensured by the algorithm, the estimators are monotone in τ. As in the calibrated finite activity model and the self-decomposable model from Chapter 2, negative jumps have a higher activity. Roughly, the intensity of small jumps is larger for short maturities while the tails are more heavy for longer maturities.

5.4. Proofs

5.4.1. Error analysis for Section 5.2

To simplify the notation we will frequently use the definition $I_h := [-\frac{1}{h}, \frac{1}{h}]$.

Drift and remainder

All estimators are constructed based on the estimator $\widehat{\psi}_n''(u) = \Delta^{-1}(\log(\varphi_{\Delta,n}(u))''$ of ψ'' from (5.1). In particular, $\widehat{\nu}_h, \widehat{N}_h$ and $\widehat{q}_{\tau,h}$ only depend on the observations via $\widehat{\psi}_n''$. As Nickl et al. (2013, Lem. 10) we first note that the drift has no effect on the estimators.

Lemma 5.15. *Let $X_k := Y_k - \Delta\gamma$ for $k = 1, \dots, n$. Then X_k are distributed according to an infinitely divisible distribution with characteristic triplet $(\sigma^2, 0, \nu)$. Denoting the*

[1] provided by the SFB 649 “Economic Risk”

estimator $\widehat{\psi}''_n$ *based on* (X_k) *and* (Y_k) *by* $\widehat{\psi}''_{X,n}$ *and* $\widehat{\psi}''_{Y,n}$, *respectively, it holds* $\widehat{\psi}''_{X,n} = \widehat{\psi}''_{X,n}$ *for all* $u \in \mathbb{R}$.

Proof. The drift causes a factor $e^{i\Delta\gamma u}$ in the empirical characteristic function $\varphi_{\Delta,n,Y}$ such that

$$\widehat{\psi}''_{n,Y}(u) = \Delta^{-1}(\log(\varphi_{\Delta,n,X}(u)) + i\Delta\gamma u)'' = \widehat{\psi}''_{n,X}(u).$$

□

Using

$$(\varphi_\Delta^{-1})' = -\Delta\psi'\varphi_\Delta^{-1}, \qquad (\varphi_\Delta^{-1})'' = -\Delta(\psi'' - \Delta(\psi')^2)\varphi_\Delta^{-1}, \tag{5.19}$$

the estimation error $\psi'' - \widehat{\psi}''_n = -\Delta^{-1}\log(\varphi_{\Delta,n}/\varphi_\Delta)''$ can be linearized similarly to Proposition 21 in Nickl et al. (2013).

Lemma 5.16. *Let* $\int x^{4+\delta}\nu(\mathrm{d}x) < \infty$ *for some* $\delta > 0$. *For* $h, \Delta \in (0,1)$ *satisfying* $n^{-1/2}(\log h^{-1})^{(1+\delta)/2}\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)} \to 0$ *as* $n \to \infty$, *it holds*

$$\begin{aligned}&\sup_{|u|\leqslant h^{-1}}\Big|\widehat{\psi}''_n(u) - \psi''(u) - \Delta^{-1}(\varphi_\Delta^{-1}(\varphi_{\Delta,n} - \varphi_\Delta))''(u)\Big| \\ =&\mathcal{O}_P\big((\Delta\|\psi'\|_{L^\infty(I_h)} + \Delta^{3/2}\|\psi'\|^2_{L^\infty(I_h)} + 1)n^{-1}\Delta^{-1/2}\log(h^{-1})^{1+\delta}\|\varphi_\Delta^{-1}\|^2_{L^\infty(I_h)}\big).\end{aligned}$$

Proof. Setting $F(y) = \log(1+y)$ and $\eta = (\varphi_{\Delta,n} - \varphi_\Delta)/\varphi_\Delta$, we use $(F\circ\eta)''(u) = F'(\eta(u))\eta''(u) + F''(\eta(u))\eta'(u)^2$ to obtain

$$|(F\circ\eta)''(u) - \eta''(u)| \lesssim \|F''\|_\infty\Big(\eta(u)\eta''(u) + \eta'(u)^2\Big). \tag{5.20}$$

On the event $\Omega_n := \{\sup_{|u|\leqslant 1/h}|(\varphi_{\Delta,n} - \varphi_\Delta)(u)/\varphi_\Delta(u)| \leqslant 1/2\}$ we thus obtain

$$\begin{aligned}&\sup_{|u|\leqslant h^{-1}}|\log(\varphi_{\Delta,n}/\varphi_\Delta)''(u) - (\varphi_\Delta^{-1}(\varphi_{\Delta,n} - \varphi_\Delta))''(u)| \\ &\qquad\lesssim \|\eta\|_{L^\infty(I_h)}\|\eta''\|_{L^\infty(I_h)} + \|\eta'\|^2_{L^\infty(I_h)}.\end{aligned}$$

To estimate $\|\eta^{(k)}\|_{L^\infty(I_h)}, k = 0,1,2$, we use (5.19) and $|\psi''(u)| \lesssim 1$ to obtain

$$\begin{aligned}\sup_{u\in I_h}\Big|(\varphi_\Delta^{-1})'(u)\Big| &\lesssim \Delta\|\psi'\|_{L^\infty(I_h)}\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)}, \\ \sup_{u\in I_h}\Big|(\varphi_\Delta^{-1})''(u)\Big| &\lesssim (\Delta^2\|\psi'\|^2_{L^\infty(I_h)} + \Delta)\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)}.\end{aligned}$$

Applying Theorem 1 by Kappus and Reiß (2010) and the moment assumption on ν, we obtain for $k = 0,1,2$ and $\delta > 0$

$$\big\|(\varphi_{\Delta,n} - \varphi_\Delta)^{(k)}\big\|_{L^\infty(I_h)} = \mathcal{O}_P(n^{-1/2}\Delta^{(k\wedge 1)/2}(\log h^{-1})^{(1+\delta)/2}).$$

This yields

$$\begin{aligned}\|\eta\|_{L^\infty(I_h)} &= \mathcal{O}_P\big(n^{-1/2}(\log h^{-1})^{(1+\delta)/2}\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)}\big), \qquad (5.21)\\ \|\eta'\|_{L^\infty(I_h)} &= \mathcal{O}_P\big((\Delta\|\psi'\|_{L^\infty(I_h)} + \Delta^{1/2})n^{-1/2}(\log h^{-1})^{(1+\delta)/2}\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)}\big), \\ \|\eta''\|_{L^\infty(I_h)} &= \mathcal{O}_P\big((\Delta^{3/2}\|\psi'\|_{L^\infty(I_h)} + \Delta^2\|\psi'\|^2_{L^\infty(I_h)} + \Delta^{1/2}) \\ &\qquad\times n^{-1/2}(\log h^{-1})^{(1+\delta)/2}\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)}\big).\end{aligned}$$

The bound (5.21) and $n^{-1/2}(\log h^{-1})^{(1+\delta)/2}\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)} \to 0$, yield $P(\Omega_n) \to 1$ which implies the assertion. $\square$

Convergence rates for distribution function estimation

Because it is a bit more difficult to derive the convergence rates for the distribution function estimator, we study this problem first. The corresponding results for $\widehat{\nu}_n$ can the be proved analogously.

We decompose the estimation error of the distribution function estimator $\widehat{N}_h$ into

$$\begin{aligned}\widehat{N}_h(t) - N(t) = &\int g_t(x)(K_h * (y^2\nu) - x^2\nu)(\mathrm{d}x)\\ &+ \int g_t(x)\,\mathcal{F}^{-1}\Big[\Big(\psi''(u) - \widehat{\psi}_n''(u)\Big)\,\mathcal{F}\,K(hu)\Big](x)\mathrm{d}x + \sigma^2\int g_t(x)K_h(x)\mathrm{d}x\\ =: &B_n(t) + S_n(t) + V_n(t),\end{aligned} \tag{5.22}$$

where B_n is the deterministic error term, S_n is the stochastic error term and V_n is the error due to the unknown volatility σ^2. The error term V_n is negligible:

Lemma 5.17. *Grant Assumption (5.6) on the kernel with order $p \geqslant s+1$. Then $V_n(t), t \neq 0$, as defined in (5.22) satisfies $|V_n(t)| \lesssim \sigma^2|t|^{-s-3}h^{s+1}$.*

Proof. We estimate

$$\begin{aligned}|V_n| \leqslant \sigma^2\int |g_t(x)K_h(x)|\mathrm{d}x \leqslant &\sigma^2\|x^{-s-1}g_t(x)\|_\infty\|x^{s+1}K_h(x)\|_{L^1}\\ = &\sigma^2|t|^{-s-3}h^{s+1}\|x^{s+1}K(x)\|_{L^1}.\end{aligned}$$

$\square$

For the bias we apply the following:

Proposition 5.18. *Suppose $\|x^2\nu\|_{L^1} < \infty$ and let $U \subseteq \mathbb{R}$ be an open set. If ν admits a Lebesgue density on U in $C^s(U)$ for some $s > 0$ and if the kernel satisfies (5.6) with order $p \geqslant s+1$, then*

$$|B_n(t)| = \Big|\int g_t(x)(K_h * (y^2\nu(\mathrm{d}y)) - x^2\nu)(\mathrm{d}x)\Big| \lesssim (|t|^{-s-4} \vee 1)h^{s+1}, \quad \textit{for all} \quad t \in U.$$

Proof. Without loss of generality let $t < 0$. Using Fubini's theorem, we rewrite

$$\int_{\mathbb{R}} g_t(x)(K_h * (y^2\nu(\mathrm{d}y)) - x^2\nu)(\mathrm{d}x) = K_h * g_t(-\bullet) * (x^2\nu)(0) - g_t(-\bullet) * (x^2\nu)(0). \tag{5.23}$$

Denoting $\overline{N}(t) = \int_{-\infty}^t x^2\nu(\mathrm{d}x), t < 0$, integration by parts yields

$$\begin{aligned}g_t(-\bullet) * (x^2\nu)(y) &= \int_{-\infty}^{t+y}\frac{x^2}{(x-y)^2}\nu(\mathrm{d}x)\\ &= t^{-2}\overline{N}(t+y) + \int_{-\infty}^0 \frac{2}{(x+t)^3}\overline{N}(x+t+y)\mathrm{d}x\\ &= t^{-2}\overline{N}(t+y) + 2\left(((\bullet + t)^{-3}\mathbb{1}_{(-\infty,0]}) * \overline{N}(t-\bullet)\right)(-y).\end{aligned}$$

From this representation we see for some sufficiently small $\delta \in (0,|t|)$ that $g_t(-\bullet) * (x^2\nu) \in C^{s+1}((-\delta/2,\delta/2))$ owing to $\overline{N}(t+\bullet) \in C^{s+1}((-\delta,\delta))$ and $(\bullet+t)^{-3}\mathbb{1}_{(-\infty,0]} \in C^{s+1}((-\delta/2,\delta/2)^c)$. The corresponding Hölder norm of the latter is of order t^{-s-4}. With a standard Taylor expansion argument and applying the order of the kernel, the approximation error is of the order $(|t|^{-s-4} \vee 1)h^{s+1}$ by (5.23). $\square$

Lemma 5.16 motivates the following definition of the linearized stochastic error term

$$L_{\Delta,n}(t) := -\Delta^{-1} \int g_t(x)\mathcal{F}^{-1}[\mathcal{F}K(h\bullet)(\varphi_\Delta^{-1}(\varphi_{\Delta,n}-\varphi_\Delta))''](x)\mathrm{d}x. \tag{5.24}$$

Using (5.19) and defining the regularized Fourier multiplier

$$m_{\Delta,h} := \frac{\mathcal{F}K(h\bullet)}{\varphi_\Delta},$$

we decompose the linearized stochastic error further into

$$\begin{aligned} L_{\Delta,n}(t) = &-\frac{1}{\Delta}\int g_t(x)\mathcal{F}^{-1}[\mathcal{F}K(h\bullet)(\varphi_\Delta^{-1}(\varphi_{\Delta,n}-\varphi_\Delta)''+2(\varphi_\Delta^{-1})'(\varphi_{\Delta,n}-\varphi_\Delta)' \\ &+(\varphi_\Delta^{-1})''(\varphi_{\Delta,n}-\varphi_\Delta))](x)\mathrm{d}x \\ =&\underbrace{-\frac{1}{\Delta}\int g_t(x)\mathcal{F}^{-1}[m_{\Delta,h}(\varphi_{\Delta,n}''-\varphi_\Delta'')](x)\mathrm{d}x}_{=:M_{\Delta,n}(t)} \\ &+2\int g_t(x)\mathcal{F}^{-1}[m_{\Delta,h}\psi'(\varphi_{\Delta,n}'-\varphi_\Delta')](x)\mathrm{d}x \\ &+\int g_t(x)\mathcal{F}^{-1}[m_{\Delta,h}(\psi''-\Delta(\psi')^2)(\varphi_{\Delta,n}-\varphi_\Delta)](x)\mathrm{d}x. \end{aligned} \tag{5.25}$$

In the following, we will refer to $M_{\Delta,n}$ as the main stochastic error term.

Proposition 5.19. *Let $U \subseteq \mathbb{R}$ be an open set, $t \in U$ and $\alpha,\beta,s,r,R>0, m>4$. Let the kernel satisfy (5.6) with $p \geqslant 1$. Then we obtain for $h,\Delta \in (0,1)$ and $\eta \in (0,1)$*

$$\sup_{(\sigma^2,\gamma,\nu)\in\mathcal{D}^s(\alpha,m,U,R)} \mathbb{E}\Big[\sup_{|t|>\eta}|L_{\Delta,n}-M_{\Delta,n}|(t)\Big] \lesssim \eta^{-2}n^{-1/2}\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)},$$

$$\sup_{(\sigma^2,\gamma,\nu)\in\mathcal{E}^s(\beta,m,U,r,R)} \mathbb{E}\Big[\sup_{|t|>\eta}|L_{\Delta,n}-M_{\Delta,n}|(t)\Big] \lesssim \frac{\log(h^{-1})+\Delta^{1/2}h^{-1}+\Delta h^{-2}}{\eta^2 n^{1/2}}e^{r\Delta h^{-\beta}}.$$

Moreover,

$$\sup_{(\sigma^2,\gamma,\nu)\in\mathcal{E}^s(\beta,m,U,r,R)} \mathbb{E}\Big[\sup_{|t|>\eta}|M_{\Delta,n}(t)|\Big] \lesssim \eta^{-2}(n\Delta)^{-1/2}\log(h^{-1})e^{r\Delta h^{-\beta}}.$$

Remark 5.20. Note that

$$\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)} \lesssim \begin{cases} h^{-\Delta\alpha+1/2} & \text{for } \Delta\alpha>1/2, \\ (\log h^{-1})^{1/2} & \text{for } \Delta\alpha=1/2, \\ 1 & \text{for } \Delta\alpha\in(0,1/2). \end{cases}$$

Proof. Due to $\|g_t\|_{L^1} = |t|^{-1}, \|g_t\|_{BV} \leqslant 2t^{-2}$, we obtain $|\mathcal{F} g_t(u)| \lesssim (t^{-1} \vee t^{-2})(1+|u|)^{-1}, u \in \mathbb{R}$. Using Plancherel's identity, we estimate for $|t| > \eta$

$$\begin{aligned}
|L_{\Delta,n} - M_{\Delta,n}|(t) \leqslant &\pi^{-1}\Big|\int \mathcal{F} g_t(-u) m_{\Delta,h}(u)\psi'(u)(\varphi_{\Delta,n} - \varphi_\Delta)'(u)\mathrm{d}u\Big| \\
&+ \frac{1}{2\pi}\Big|\int \mathcal{F} g_t(-u) m_{\Delta,h}(u)(\psi''(u) - \Delta\psi'(u)^2)(\varphi_{\Delta,n} - \varphi_\Delta)(u)\mathrm{d}u\Big| \\
\lesssim &\eta^{-2}\int (1+|u|)^{-1}|m_{\Delta,h}(u)\psi'(u)(\varphi_{\Delta,n} - \varphi_\Delta)'(u)|\mathrm{d}u \\
&+ \eta^{-2}\int (1+|u|)^{-1}|m_{\Delta,h}(u)\psi''(u)(\varphi_{\Delta,n} - \varphi_\Delta)(u)|\mathrm{d}u \\
&+ \eta^{-2}\Delta\int (1+|u|)^{-1}|m_{\Delta,h}(u)\psi'(u)^2(\varphi_{\Delta,n} - \varphi_\Delta)(u)|\mathrm{d}u.
\end{aligned}$$

Due to $\mathbb{E}[Y_1^{2l}] \lesssim \Delta^{l\wedge 1}, l = 0,1,2$, we have, moreover,

$$\sup_{u\in\mathbb{R}} \mathbb{E}[(\varphi_{\Delta,n}^{(l)} - \varphi_\Delta^{(l)})^2(u)] \lesssim n^{-1}\Delta^{(l\wedge 1)}, \quad \text{for } l = 0,1,2. \tag{5.26}$$

Fubini's theorem and Jensen's inequality yield then

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{|t|>\eta} |L_{\Delta,n} - M_{\Delta,n}|(t)\Big] \\
\lesssim &\eta^{-2}\Big(\int (1+|u|)^{-1}|m_{\Delta,h}(u)\psi'(u)|\,\mathbb{E}\left[(\varphi'_{\Delta,n} - \varphi'_\Delta)^2(u)\right]^{1/2}\mathrm{d}u \\
&+ \int (1+|u|)^{-1}|m_{\Delta,h}(u)\psi''(u)|\,\mathbb{E}[(\varphi_{\Delta,n} - \varphi_\Delta)^2(u)]^{1/2}\mathrm{d}u \\
&+ \Delta\int (1+|u|)^{-1}|m_{\Delta,h}(u)\psi'(u)^2|\,\mathbb{E}[(\varphi_{\Delta,n} - \varphi_\Delta)^2(u)]^{1/2}\mathrm{d}u\Big) \\
\lesssim &n^{-1/2}\eta^{-2}\Big(\Delta^{1/2}\big\|(1+|u|)^{-1}m_{\Delta,h}(u)\psi'(u)\big\|_{L^1} + \big\|(1+|u|)^{-1}m_{\Delta,h}(u)\psi''(u)\big\|_{L^1} \\
&+ \Delta\big\|(1+|u|)^{-1}m_{\Delta,h}(u)\psi'(u)^2\big\|_{L^1}\Big).
\end{aligned} \tag{5.27}$$

To deal with $(\sigma^2, \gamma, \nu) \in \mathcal{D}^s(\alpha, m, U, R)$, we note that the assumptions $\|x^4\nu\|_{L^1} < \infty$ and $\|x\nu\|_\infty < \infty$ imply $x\nu, x^2\nu \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Moreover, $\sigma^2 = 0$ and by Lemma 5.15 we can assume $\gamma_0 = \gamma - \int_{-1}^1 x\mathrm{d}\nu = 0$, cf. (2.3). Hence, $i\psi' = -\mathcal{F}[x\nu]$ and $\psi'' = -\mathcal{F}[x^2\nu]$. Therefore, we estimate (5.27) with use of the Cauchy–Schwarz inequality

$$\begin{aligned}
&\mathbb{E}\,[\sup_{|t|>\eta} |L_{\Delta,n} - M_{\Delta,n}|(t)] \\
\lesssim &n^{-1/2}\eta^{-2}\Big(\Delta^{1/2}\|(1+|u|)^{-1}m_{\Delta,h}(u)\|_{L^2}\|\psi'\|_{L^2} + \|(1+|u|)^{-1}m_{\Delta,h}(u)\|_{L^2}\|\psi''\|_{L^2} \\
&+ \Delta\|(1+|u|)^{-1}m_{\Delta,h}(u)\|_{L^2}\|\psi'\|_{L^2}\|\psi'\|_\infty\Big) \\
\lesssim &n^{-1/2}\eta^{-2}(\Delta^{1/2}\|x\nu\|_{L^2} + \|x^2\nu\|_{L^2} + \Delta\|x\nu\|_{L^2}\|x\nu\|_{L^1})\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)},
\end{aligned} \tag{5.28}$$

which yields the assertion for $\mathcal{D}^s(\alpha, m, U, R)$.

Now we consider the case $(\sigma^2,\gamma,\nu)\in\mathcal{E}^s(\beta,m,U,r,R)$. The exponential decay of φ_Δ, the properties of K, $|\psi'(u)|\lesssim 1+|u|$ and $|\psi''(u)|\lesssim 1$ yield

$$\begin{aligned}\Delta^{1/2}\big\|(1+|u|)^{-1}m_{\Delta,h}(u)\psi'(u)\big\|_{L^1} &\lesssim \Delta^{1/2}\big\|m_{\Delta,h}\big\|_{L^1}\\ &\lesssim \Delta^{1/2}\int_{-1/h}^{1/h} e^{r\Delta|u|^\beta}\mathrm{d}u \lesssim \Delta^{1/2}h^{-1}\exp\big(r\Delta h^{-\beta}\big),\\ \big\|(1+|u|)^{-1}m_{\Delta,h}(u)\psi''(u)\big\|_{L^1} &\lesssim \big\|(1+u)^{-1}m_{\Delta,h}(u)\big\|_{L^1}\lesssim \log(h^{-1})\exp\big(r\Delta h^{-\beta}\big),\\ \Delta\big\|(1+|u|)^{-1}m_{\Delta,h}(u)\psi'(u)^2]\big\|_{L^1} &\lesssim \Delta\big\|(1+u)m_{\Delta,h}(u)\big\|_{L^1}\lesssim \Delta h^{-2}\exp\big(r\Delta h^{-\beta}\big).\end{aligned}$$

We conclude the claimed estimate for $|L_{\Delta,n}-M_{\Delta,n}|$ in $\mathcal{E}^s(\beta,m,U,r,R)$ by plugging these estimates into (5.27). Similarly, we estimate

$$\begin{aligned}\mathbb{E}\Big[\sup_{|t|>\eta}|M_{\Delta,n}|(t)\Big] &\lesssim \eta^{-2}\Delta^{-1}\,\mathbb{E}\Big[\int(1+|u|)^{-1}m_{\Delta,h}(u)(\varphi''_{\Delta,n}-\varphi''_\Delta)(u)\mathrm{d}u\Big]\\ &=\eta^{-2}(n\Delta)^{-1/2}\|(1+|u|)^{-1}m_{\Delta,h}(u)\|_{L^1}\\ &\lesssim \eta^{-2}(n\Delta)^{-1/2}\log(h^{-1})\exp\big(r\Delta h^{-\beta}\big).\end{aligned}$$ □

For the main stochastic error term in the mildly ill-posed case we will need the following concentration result:

Proposition 5.21. *Let $U\subseteq\mathbb{R}$ be an open set, $t\in U$ and $\alpha,s,R>0,m>4$ and let the kernel satisfy (5.6) with $p\geqslant 1$. Then there is some $c>0$ such that for any $h,\Delta,\eta\in(0,1)$ and $\kappa_0>0$*

$$\begin{aligned}\sup_{(\sigma^2,\gamma,\nu)\in\mathcal{D}^s(\alpha,m,U,R)}\ \sup_{|t|>\eta}P\Big(|M_{\Delta,n}(t)|>\kappa_0(\eta^{-1}\vee 1)(n\Delta)^{-1/2}\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}\Big)\\ \leqslant 2\exp\Big(-\frac{c\kappa_0^2}{1+\kappa_0(\eta^{-1}\vee 1)(hn\Delta)^{-1/2}}\Big).\end{aligned}$$

Proof. We represent $M_{\Delta,n}(t)$ as sum of i.i.d. random variables via

$$M_{\Delta,n}(t)=\sum_{k=1}^n(\xi_k(t)-\mathbb{E}[\xi_k(t)]),\quad \xi_k(t):=(n\Delta)^{-1}\int g_t(x)\,\mathcal{F}^{-1}\big[m_{\Delta,h}(u)Y_k^2e^{iuY_k}\big](x)\mathrm{d}x.$$

Applying Plancherel's identity, $\xi_k(t)$ can be rewritten as

$$\xi_k(t)=\frac{1}{2\pi n\Delta}Y_k^2\int\mathcal{F}g_t(-u)m_{\Delta,h}(u)e^{iuY_k}\mathrm{d}u=\frac{1}{n\Delta}Y_k^2\,\mathcal{F}^{-1}[\mathcal{F}g_t(-\bullet)m_{\Delta,h}](-Y_k).$$

To estimate $\mathrm{Var}(\xi_k)\leqslant\mathbb{E}[\xi_k^2]$, we use $g_t\in BV(\mathbb{R})$ to decompose $g_t=g_t^s+g_t^c$ into a singular component and a continuous component satisfying for $r>\Delta\alpha$

$$\begin{aligned}\max\Big\{\mathcal{F}g_t^s(u),\mathcal{F}[xg_t^s](u),\mathcal{F}[x^2g_t^s](u)\Big\}&\lesssim(t^{-2}\vee 1)(1+|u|)^{-1},\\ \max\Big\{\sup_{|t|>\eta}\|g_t^c\|_{C^r},\sup_{|t|>\eta}\|x^2g_t^c\|_{C^r}\Big\}&\lesssim 1.\end{aligned}$$

This allows to decompose

$$\mathbb{E}[\xi_k^2] \leqslant \frac{2}{(n\Delta)^2}\Big(\mathbb{E}\Big[Y_1^4\,\mathcal{F}^{-1}\left[\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h}\right](-Y_1)^2\Big] \\ + \mathbb{E}\Big[Y_1^4\,\mathcal{F}^{-1}\left[\mathcal{F}\,g_t^c(-\bullet)m_{\Delta,h}\right](-Y_1)^2\Big]\Big) =: \frac{2}{(n\Delta)^2}(E_s+E_c). \tag{5.29}$$

To estimate E_c in (5.29), we apply the Fourier multiplier Theorem 5.5 to see that

$$\|\,\mathcal{F}^{-1}\left[\mathcal{F}\,g_t^c(-\bullet)m_{\Delta,h}\right]\|_\infty \lesssim \|g_t^c * K_h\|_{C^r} \leqslant \|K\|_{L^1}\|g_t^c\|_{C^r}$$

for any $r > \Delta\alpha$. Consequently, $(n\Delta)^{-2}E_s \lesssim (n\Delta)^{-2}\,\mathbb{E}[Y_1^4] \lesssim n^{-2}\Delta^{-1}$ because Y_1 has fourth moments due to $\|x^4\nu\|_{L^1} < \infty$.

It remains to bound E_s from (5.29). Using again $\gamma_0 = 0$ by Lemma 5.15, we infer from $\mathcal{F}[ixP_\Delta] = \varphi_\Delta' = \Delta\psi'\varphi = \Delta\,\mathcal{F}[ix\nu]\varphi_\Delta$ that

$$xP_\Delta = \Delta(x\nu) * P_\Delta \tag{5.30}$$

and thus xP_Δ has a bounded density satisfying $\|xP_\Delta\|_\infty \leqslant \Delta\|x\nu\|_\infty$. Together with the Cauchy–Schwarz inequality and Plancherel's identity we obtain

$$\begin{aligned} E_s \leqslant&\Delta\|x\nu\|_\infty \int |y|^3\Big(\mathcal{F}^{-1}\left[\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h}\right](-y)\Big)^2\mathrm{d}y \\ =&\Delta\|x\nu\|_\infty \int \Big|\mathcal{F}^{-1}\left[(\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h})''\right](-y)\,\mathcal{F}^{-1}\left[(\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h})'\right](-y)\Big|\mathrm{d}y \\ \leqslant&\Delta\|x\nu\|_\infty \left\|(\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h})''\right\|_{L^2}\left\|(\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h})'\right\|_{L^2}. \end{aligned} \tag{5.31}$$

The derivatives of the regularized Fourier multiplier are given by

$$m_{\Delta,h}'(u) = ih\frac{\mathcal{F}[xK](hu)}{\varphi_\Delta(u)} - \Delta\psi'(u)m_{\Delta,h}(u) = ih\frac{\mathcal{F}[xK](hu)}{\varphi_\Delta(u)} - i\Delta\,\mathcal{F}[x\nu](u)m_{\Delta,h}(u), \tag{5.32}$$

$$m_{\Delta,h}''(u) = -h^2\frac{\mathcal{F}[x^2K](hu)}{\varphi_\Delta(u)} - 2i\Delta h\psi'(u)\frac{\mathcal{F}[xK](hu)}{\varphi_\Delta(u)} \\ - \Delta(\psi''(u)(u) + \Delta\psi'(u)^2)m_{\Delta,h}(u). \tag{5.33}$$

To bound the L^2-norms in (5.31), we use the properties of K, the decay assumption on φ_Δ and $\psi',\psi'' \in L^\infty(\mathbb{R})$ to obtain

$$\begin{aligned} \|(\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h})'\|_{L^2} \leqslant&\|\,\mathcal{F}[xg_t^s](-\bullet)m_{\Delta,h}\|_{L^2} + \|\,\mathcal{F}[g_t^s](-\bullet)m_{\Delta,h}'\|_{L^2} \\ \lesssim&(t^{-2}\vee 1)(1+h+\Delta)\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}, \\ \|(\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h})''\|_{L^2} \leqslant&\|\,\mathcal{F}[x^2g_t^s](-\bullet)m_{\Delta,h}\|_{L^2} \\ &+ 2\|\,\mathcal{F}[xg_t^s](-\bullet)m_{\Delta,h}'\|_{L^2} + \|\,\mathcal{F}\,g_t^s(-\bullet)m_{\Delta,h}''\|_{L^2} \\ \lesssim&(t^{-2}\vee 1)(1+h+\Delta+h^2+\Delta h)\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}. \end{aligned} \tag{5.34}$$

Therefore, $(\Delta n)^{-2}E_s \lesssim (t^{-2}\vee 1)n^{-2}\Delta^{-1}\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}^2$ which implies

$$\operatorname{Var}(\xi_k(t)) \lesssim \frac{t^{-2}\vee 1}{n^2\Delta}\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}^2. \tag{5.35}$$

Using (5.32), (5.33), $x\nu \in L^2(\mathbb{R})$, $|\mathcal{F} g_t(u)| \lesssim (t^{-2} \vee 1)(1+|u|)^{-1}$ and $\|xg_t\|_{L^2} \lesssim |t|^{-1}$ we deterministically bound $\xi_k(t)$ by

$$\begin{aligned}
|\xi_k(t)| \leqslant &(n\Delta)^{-1}\| \mathcal{F}^{-1}[(\mathcal{F} g_t(-\bullet)m_{\Delta,h})'']\|_\infty \\
\leqslant &(n\Delta)^{-1}\Big(\| \mathcal{F}^{-1}\left[\mathcal{F}[x^2 g_t](-\bullet)m_{\Delta,h}\right]\|_\infty \\
&+ 2\|\mathcal{F}[xg_t](-\bullet)m'_{\Delta,h}\|_{L^1} + \|\mathcal{F} g_t(-\bullet)m''_{\Delta,h}\|_{L^1}\Big) \\
\lesssim &(n\Delta)^{-1}\Big(\| \mathcal{F}^{-1}\left[\mathcal{F}[x^2 g_t](-\bullet)m_{\Delta,h}\right]\|_\infty + \|xg_t\|_{L^2}\|m'_{\Delta,h}\|_{L^2} \\
&+ \|(1+|u|)\mathcal{F} g_t(u)\|_\infty \|(1+|u|)^{-1}m''_{\Delta,h}(u)\|_{L^1}\Big) \\
\lesssim &(n\Delta)^{-1}\Big(\|\mathcal{F}[x^2 g_t^s](-\bullet)m_{\Delta,h}\|_{L^1} + \| \mathcal{F}^{-1}\left[\mathcal{F}[x^2 g_t^c](-\bullet)m_{\Delta,h}\right]\|_\infty \\
&+ \|xg_t\|_{L^2}(h\|(1+|u|)^{\Delta\alpha}\|_{L^2(I_h)} + \Delta h^{-\Delta\alpha}\|x\nu\|_{L^2}) \\
&+ (t^{-2} \vee 1)(\Delta + h^2 + \Delta h)\|(1+|u|)^{\Delta\alpha-1}\|_{L^1(I_h)}\Big) \\
\lesssim &(n\Delta)^{-1}\|\mathcal{F}[x^2 g_t^s](-\bullet)m_{\Delta,h}\|_{L^1} + (n\Delta)^{-1}\| \mathcal{F}^{-1}\left[\mathcal{F}[x^2 g_t^c](-\bullet)m_{\Delta,h}\right]\|_\infty \\
&+ (n\Delta)^{-1}(t^{-2} \vee 1)h^{-1/2}\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}.
\end{aligned}$$

The term corresponding to the singular part g_t^s in the above display can be estimated by

$$\begin{aligned}
(n\Delta)^{-1}\|\mathcal{F}[x^2 g_t^s](-\bullet)m_{\Delta,h}\|_{L^1} &\lesssim (n\Delta)^{-1}\|(1+|u|)\mathcal{F}[x^2 g_t^s](u)\|_\infty \|(1+|u|)^{\Delta\alpha-1}\|_{L^1(I_h)} \\
&\lesssim (n\Delta)^{-1}(t^{-2} \vee 1)h^{-1/2}\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}.
\end{aligned}$$

For the continuous part g_t^c we apply the Fourier multiplier theorem as above to see that $\| \mathcal{F}^{-1}\left[\mathcal{F}[x^2 g_t^c](-\bullet)m_{\Delta,h}\right]\|_\infty \lesssim \|x^2 g_t^c\|_{C^r}$ for any $r > \Delta\alpha$. Therefore,

$$\begin{aligned}
|\xi_k(t)| \leqslant &(n\Delta)^{-1}\| \mathcal{F}^{-1}[(\mathcal{F} g_t(-\bullet)m_{\Delta,h})'']\|_\infty \\
\lesssim &(n\Delta)^{-1}(t^{-2} \vee 1)h^{-1/2}\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}. \qquad (5.36)
\end{aligned}$$

Using (5.35) and (5.36), Bernstein's inequality yields for some constant $c > 0$ the claimed concentration result. □

Combining the previous results, we obtain minimax convergence rates for estimating the (generalized) distribution function of the jump measure.

Proof of Proposition 5.7. In the following, t is fixed and thus omitted in the constants. Using the error decomposition (5.22), Lemma 5.17, Proposition 5.18, we obtain

$$\begin{aligned}
|\widehat{N}_h(t) - N(t)| \leqslant &|B_n(t)| + |S_n(t)| + |V_n(t)| \\
\lesssim &h^{s+1} + |S_n(t)|.
\end{aligned}$$

Using $|\mathcal{F} g_t(u)| \lesssim (1+|u|)^{-1}$, Plancherel's identity and Lemma 5.16 yield for the stochastic error from (5.22) and the linearized stochastic error term $L_{\Delta,n}$ defined in (5.24)

$$\begin{aligned}
&\big|S_n(t) - L_{\Delta,n}(t)\big| \\
=&\Big|\int g_t(x)\,\mathcal{F}^{-1}\Big[\mathcal{F} K(hu)\Big(\widehat{\psi}_n''(u) - \psi''(u) - \Delta^{-1}\big(\varphi_\Delta^{-1}(\varphi_{\Delta,n} - \varphi_\Delta)\big)''(u)\Big)\Big](x)\mathrm{d}x\Big| \\
\lesssim& \sup_{|u|\leqslant h^{-1}} \big|\widehat{\psi}_n''(u) - \psi''(u) - \Delta^{-1}\big(\varphi_\Delta^{-1}(\varphi_{\Delta,n} - \varphi_\Delta)\big)''(u)\big| \int (1+|u|)^{-1}|\mathcal{F} K(hu)|\mathrm{d}u \\
\leqslant& \sup_{|u|\leqslant h^{-1}} \big|\widehat{\psi}_n''(u) - \psi''(u) - \Delta^{-1}\big(\varphi_\Delta^{-1}(\varphi_{\Delta,n} - \varphi_\Delta)\big)''(u)\big|\|K\|_{L^1}\int_{-1/h}^{1/h}(1+|u|)^{-1}\mathrm{d}u \\
=&\mathcal{O}_P\big((\Delta\|\psi'\|_{L^\infty(I_h)} + \Delta^{3/2}\|\psi'\|^2_{L^\infty(I_h)} + 1)\frac{|\log h|^{2+\delta}}{n\Delta^{1/2}}\|\varphi_\Delta^{-1}\|^2_{L^\infty(I_h)}\big),
\end{aligned} \tag{5.37}$$

provided that $n^{-1/2}(\log h^{-1})^{(1+\delta)/2}\|\varphi_\Delta^{-1}\|_{L^\infty(I_h)} \to 0$. The latter condition is satisfied for the choices $h = h_{n,\Delta}$ in both cases.

Let $(\sigma^2,\gamma,\nu) \in \mathcal{D}^s(\alpha, m, U, R)$. We conclude from (5.37), where ψ' is uniformly bounded and φ_Δ decays polynomially, and Propositions 5.19 and 5.21 for $\Delta\alpha > 1/2$

$$\begin{aligned}
S_n(t) =&L_{\Delta,n}(t) + \mathcal{O}_P\big(n^{-1}\Delta^{-1/2}\log(h^{-1})^{2+\delta}h^{-2\Delta\alpha}\big) \\
=&\mathcal{O}_P\big((n\Delta)^{-1/2}h^{-\Delta\alpha+1/2} + n^{-1}\Delta^{-1/2}\log(h^{-1})^{2+\delta}h^{-2\Delta\alpha}\big) \\
=&\mathcal{O}_P\big((n\Delta)^{-1/2}h^{-\Delta\alpha+1/2}(1 + n^{-1/2}\log(h^{-1})^{2+\delta}h^{-\Delta\alpha-1/2})\big).
\end{aligned}$$

Therefore, we obtain for $h = h_{n,\Delta} = (n\Delta)^{-1/(2s+2\Delta\alpha+1)}$

$$\begin{aligned}
&\big|\widehat{N}_h(t) - N(t)\big| \\
=&\mathcal{O}_P\Big((n\Delta)^{-(s+1)/(2s+2\Delta\alpha+1)}\big(1 + \log(n\Delta)^{2+\delta}n^{-s/(2s+2\Delta\alpha+1)}\Delta^{(\Delta\alpha+1/2)/(2s+2\Delta\alpha+1)}\big)\Big).
\end{aligned}$$

For $\Delta\alpha \leqslant 1/2$ the rate $r_{n,\Delta}$ follows similarly.

Let us consider the case $(\sigma^2,\gamma,\nu) \in \mathcal{E}^s(\beta, m, U, r, R)$. Owing to $\|\psi'\|_{L^\infty(I_h)} \lesssim h^{-1}$ and the exponential decay of φ_Δ, we obtain from (5.37) and Proposition 5.19 that

$$\begin{aligned}
S_n(t) =&\mathcal{O}_P\big((n\Delta)^{-1/2}\big(\log(h^{-1}) + \Delta h^{-1} + \Delta^{3/2}h^{-2} \\
&+ n^{-1/2}\log(h^{-1})^{2+\delta}(\Delta h^{-1} + \Delta^{3/2}h^{-2} + 1)\big)\exp(r\Delta h^{-\beta})\big).
\end{aligned}$$

Therefore, plugging in $h = h_{n,\Delta} = (\frac{\delta}{2r}\frac{\log(n\Delta)}{\Delta})^{-1/\beta}, \delta \in (0,1)$, yields

$$\begin{aligned}
&\big|\widehat{N}_h(t) - N(t)\big| \\
=&\mathcal{O}_P\Big(\Big(\frac{\log(n\Delta)}{\Delta}\Big)^{-(s+1)/\beta} + (n\Delta)^{-(1-\delta)/2}\log(n\Delta)^{2/\beta}(|\log\Delta| + \Delta^{1-1/\beta} + \Delta^{3/2-2/\beta})\Big).
\end{aligned}$$

□

Uniform loss for density estimation

Applying a similar decomposition as in (5.22) and the linearization Lemma 5.16, we obtain

$$\begin{aligned}
&\widehat{\nu}_h(t)-\nu(t)\\
=&\frac{1}{t^2}\Big((K_h*(y^2\nu)-x^2\nu)(t)+\mathcal{F}^{-1}\Big[\Big(\psi''(u)-\widehat{\psi}_n''(u)\Big)\mathcal{F}K(hu)\Big](t)+\sigma^2K_h(t)\Big) \quad (5.38)\\
=&\frac{1}{t^2}\Big(\underbrace{(K_h*(y^2\nu)-x^2\nu)(t)}_{=:B^\nu(t)}\underbrace{-\frac{1}{\Delta}\mathcal{F}^{-1}\Big[\mathcal{F}K(hu)\Big(\frac{\varphi_{\Delta,n}-\varphi_\Delta}{\varphi_\Delta}\Big)''(u)\Big](t)}_{=:L^\nu_{\Delta,n}(t)}\Big)\\
&+\frac{1}{t^2}(R_{\Delta,n}+\sigma^2K_h(t))
\end{aligned}$$

for some remainder $R_{\Delta,n}$ which is of order

$$\begin{aligned}
|R_{\Delta,n}|=\mathcal{O}_P\big((\Delta\|\psi'\|_{L^\infty(I_h)}+\Delta^{3/2}\|\psi'\|^2_{L^\infty(I_h)}+1) \quad (5.39)\\
\times n^{-1}\Delta^{-1/2}h^{-1}\log(h^{-1})^{1+\delta}\|\varphi_\Delta^{-1}\|^2_{L^\infty(I_h)}\big).
\end{aligned}$$

We will need the following concentration result for the main stochastic error term of the density estimation problem

$$M^\nu_{\Delta,n}(t)=-\frac{1}{n\Delta}\sum_{k=1}^n\mathcal{F}^{-1}\Big[m_{\Delta,h}(Y_k^2e^{iuY_k}-\mathbb{E}[Y_k^2e^{iuY_k}])\Big](t), \quad (5.40)$$

where we recall $m_{\Delta,h}=\mathcal{F}K(h\bullet)/\varphi_\Delta$. We will prove it analogously to Proposition 5.21.

Lemma 5.22. *Let $\alpha,R>0,m>4,U\subseteq\mathbb{R}$ and $t\neq 0$ and let the kernel satisfy (5.6) for $p\geqslant 1$. If $(\sigma^2,\gamma,\nu)\in\mathcal{D}^s(\alpha,m,U,R)$, then there is some constant $c>0$, depending only on α,R, such that for any $\kappa_0>0$ and any $n\in\mathbb{N},\Delta,h>0$*

$$P\Big(|M^\nu_{\Delta,n}(t)|>\kappa_0(\Delta n)^{-1/2}h^{-\Delta\alpha-1/2}\Big)\leqslant 2\exp\left(-\frac{c\kappa_0^2}{(1+t^3)(1+\kappa_0(n\Delta h)^{-1/2})}\right).$$

Proof. We apply Bernstein's inequality to the sum of the independent and centered random variables

$$M^\nu_{\Delta,n}(t)=-\sum_{k=1}^n(\xi_k-\mathbb{E}[\xi_k]),\quad \text{with}\quad \xi_k:=\frac{1}{n\Delta}\mathcal{F}^{-1}\Big[m_{\Delta,h}(u)Y_k^2e^{iuY_k}\Big](t).$$

$\operatorname{Var}(\xi_k)$ can be estimated similarly to (5.31). We obtain by (5.30), the Cauchy–Schwarz inequality, Plancherel's identity, (5.32) and (5.33)

$$\begin{aligned}
\operatorname{Var}(\xi_k)\leqslant&\,\mathbb{E}\left[\xi_k^2\right]=(n\Delta)^{-2}\,\mathbb{E}\Big[Y_k^4\Big(\mathcal{F}^{-1}\Big[m_{\Delta,h}(u)e^{-iut}\Big](-Y_k)\Big)^2\Big]\\
\leqslant&\frac{1}{n^2\Delta}\|x\nu\|_\infty\|y^2\mathcal{F}^{-1}\Big[m_{\Delta,h}(u)e^{-iut}\Big](-y)\|_{L^2}\|y\,\mathcal{F}^{-1}\Big[m_{\Delta,h}(u)e^{-iut}\Big](-y)\|_{L^2}\\
=&\frac{1}{n^2\Delta}\frac{\|x\nu\|_\infty}{2\pi}\|(m_{\Delta,h}(u)e^{-iut})''\|_{L^2}\|(m_{\Delta,h}(u)e^{-iut})'\|_{L^2}\\
\lesssim&\frac{1}{n^2\Delta}(1+t^3)\|(1+|u|)^{\Delta\alpha}\|^2_{L^2(I_h)}\lesssim\frac{1}{n^2\Delta}(1+t^3)h^{-2\Delta\alpha-1}.
\end{aligned}$$

Moreover, ξ_k admits the deterministic bound

$$\begin{aligned}
|\xi_k| =& \frac{1}{n\Delta} Y_k^2 | \mathcal{F}^{-1} \Big[m_{\Delta,h}(u) e^{-iut} \Big](-Y_k)| \\
\leqslant& \frac{1}{n\Delta} | \mathcal{F}^{-1} \left[(m_{\Delta,h}(u) e^{-iut})'' \right](-Y_k)| \\
\leqslant& \frac{1}{2\pi n\Delta} \left(\|m''_{\Delta,h}\|_{L^1} + 2t\|m'_{\Delta,h}\|_{L^1} + t^2 \|m_{\Delta,h}\|_{L^1} \right) \\
\lesssim& \frac{1}{n\Delta} (1+t^2) \|(1+|u|)^{\Delta\alpha}\|_{L^1(I_h)} \lesssim \frac{1+t^2}{n\Delta} h^{-\Delta\alpha-1}.
\end{aligned}$$

Therefore, Bernstein's inequality yields for a constant $c > 0$ and any $\kappa > 0$

$$P\left(|M^{\nu}_{\Delta,n}(t)| > \kappa \right) \leqslant 2 \exp\left(- \frac{c\Delta n \kappa^2}{(1+t^3) h^{-2\Delta\alpha-1} + \kappa(1+t^2) h^{-\Delta\alpha-1}} \right).$$

Choosing $\kappa = \kappa_0 (\Delta n)^{-1/2} h^{-\Delta\alpha-1/2}$ for $\kappa_0 > 0$, we conclude

$$P\left(|M^{\nu}_{\Delta,n}(t)| > \kappa_0 (\Delta n)^{-1/2} h^{-\Delta\alpha-1/2} \right) \leqslant 2 \exp\left(- \frac{c\kappa_0^2}{(1+t^3)(1+\kappa_0 (n\Delta h)^{-1/2})} \right). \quad \square$$

Proposition 5.6 is an immediate consequence of the following result.

Proposition 5.23. *Let $\alpha, \beta, s, r, R > 0, m > 4$ and let the kernel satisfy (5.6) with order $p \geqslant s$ and let $U \subseteq \mathbb{R}$ be a bounded, open set which is bounded away from zero. Then we have*

(i) uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{D}^s(\alpha, m, U, R)$, if $\log(n\Delta)/(n\Delta h) \to 0$,

$$\sup_{t\in U} |\widehat{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{D}^s}\Big(h^s + (\frac{\log n\Delta}{n\Delta})^{1/2} h^{-\Delta\alpha-1/2} \Big),$$

(ii) uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{E}^s(\beta, m, U, r, R)$

$$\sup_{t\in U} |\widehat{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{E}^s}\Big(h^s + (n\Delta)^{-1/2}(h^{-1} + \Delta h^{-2} + \Delta^{3/2} h^{-3}) e^{r\Delta h^{-\beta}} \Big).$$

Proof. We start with the error decomposition (5.38). By standard approximation arguments the deterministic error satisfies $\sup_{t\in U} |B^{\nu}(t)| \lesssim h^s$ if $\nu \in C^s(U)$ for an open set U , $s > 0$ and if the kernel satisfies (5.6) with order $p \geqslant s$. Moreover, the assumptions $\operatorname{supp} \mathcal{F} K \subseteq [-1,1]$ and $x^{p+1}K(x) \in L^1(\mathbb{R})$ imply $\|x^{p+1}K(x)\|_\infty < \infty$ which yields

$$|\sigma^2 K_h(t)| \leqslant \sigma^2 h^{-1} \sup_{|x|>|t|/h} |K(x)| \lesssim \sigma^2 |t|^{-s-1} h^s.$$

Since U is bounded away from zero, the previous display gives a uniform bound on U. Using the main stochastic error term $M^{\nu}_{\Delta,n}$ from (5.40), the linearized stochastic error term can be decomposed similarly to (5.25) into

$$\begin{aligned}
L^{\nu}_{\Delta,n}(t) =& M^{\nu}_{\Delta,n}(t) + 2\,\mathcal{F}^{-1}\left[m_{\Delta,h} \psi' (\varphi_{\Delta,n} - \varphi_\Delta)' \right](t) \\
&+ \mathcal{F}^{-1}[m_{\Delta,h}(\psi'' - \Delta(\psi')^2)(\varphi_{\Delta,n} - \varphi_\Delta)](t). \qquad (5.41)
\end{aligned}$$

To derive the appropriate bounds for $R_{\Delta,n}$ and $L^{\nu}_{\Delta,n}$, we will distinguish again between the mildly and the severely ill-posed case.

We start with the severely ill-posed case $(\sigma^2,\gamma,\nu)\in\mathcal{E}^s(\beta,m,U,r,R)$. Using Fubini's theorem and the properties of $m_{\Delta,h}$ as well as $|\psi''(u)|\lesssim 1, |\psi'(u)|\lesssim 1+|u|, u\in\mathbb{R}$, and (5.26), we obtain

$$\begin{aligned}
&\mathbb{E}[\sup_{t\in U}|L^{\nu}_{\Delta,n}(t)|]\\
&\quad\leqslant \Delta^{-1}\,\mathbb{E}\Big[\|\mathcal{F}^{-1}\Big[m_{\Delta,h}(\varphi''_{\Delta,n}-\varphi''_{\Delta})\|_\infty+2\,\mathbb{E}\Big[\|\,\mathcal{F}^{-1}\left[m_{\Delta,h}\psi'(\varphi_{\Delta,n}-\varphi_{\Delta})'\right]\|_\infty\Big]\\
&\qquad+\mathbb{E}\Big[\|\mathcal{F}^{-1}[m_{\Delta,h}(\psi''-\Delta(\psi')^2)(\varphi_{\Delta,n}-\varphi_{\Delta})]\|_\infty\Big]\\
&\quad\lesssim\int_{-1/h}^{1/h}\Big(\Delta^{-1}\,\mathbb{E}[(\varphi''_{\Delta,n}(u)-\varphi''_{\Delta}(u))^2]^{1/2}+\mathbb{E}[(\varphi'_{\Delta,n}(u)-\varphi'_{\Delta}(u))^2]^{1/2}(1+|u|)\\
&\qquad+\mathbb{E}[(\varphi_{\Delta,n}(u)-\varphi_{\Delta}(u))^2]^{1/2}(1+\Delta(1+|u|)^2)\Big)\exp(r\Delta|u|^\beta)\mathrm{d}u\\
&\quad\leqslant(n\Delta)^{-1/2}\int_{-1/h}^{1/h}\Big(1+\Delta(1+|u|)+\Delta^{1/2}+\Delta^{3/2}(1+|u|)^2\Big)\exp(r\Delta|u|^\beta)\mathrm{d}u\\
&\quad\lesssim(n\Delta)^{-1/2}(h^{-1}+\Delta h^{-2}+\Delta^{3/2}h^{-3})\exp(r\Delta h^{-\beta}).
\end{aligned}$$

The remainder (5.39) is of smaller order

$$|R_{\Delta,n}|=\mathcal{O}_P\big(n^{-1}\Delta^{-1/2}\log(h^{-1})^{1+\delta}(h^{-1}+\Delta h^{-2}+\Delta^{3/2}h^{-3}))\exp(2r\Delta h^{-\beta})\Big).$$

Now let us consider $(\sigma^2,\gamma,\nu)\in\mathcal{D}^s(\alpha,m,U,R)$, where we have

$$|R_{\Delta,n}|=\mathcal{O}_P((\Delta+\Delta^{3/2}+1)n^{-1}\Delta^{-1/2}\log(h^{-1})^{1+\delta}h^{-2\Delta\alpha-1}).$$

To bound $\sup_{t\in U}|L^{\nu}_{\Delta,n}(t)|$, we note for the second and the third term in (5.41) that similarly to (5.28) with the Cauchy–Schwarz inequality and Fubini's theorem

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in U}|L^{\nu}_{\Delta,n}(t)-M^{\nu}_{\Delta,n}(t)|\Big]\\
\leqslant&2\,\mathbb{E}\Big[\|m_{\Delta,h}\psi'(\varphi_{\Delta,n}-\varphi_{\Delta})'\|_{L^1}\Big]+\mathbb{E}\Big[\|m_{\Delta,h}(\psi''-\Delta(\psi')^2)(\varphi_{\Delta,n}-\varphi_{\Delta})\|_{L^1}\Big]\\
\lesssim&2\|x\nu\|_{L^2}\,\mathbb{E}\left[\|m_{\Delta,h}(\varphi_{\Delta,n}-\varphi_{\Delta})'\|_{L^2}\right]+\|x^2\nu\|_{L^2}\,\mathbb{E}\left[\|m_{\Delta,h}(\varphi_{\Delta,n}-\varphi_{\Delta})\|_{L^2}\right]\\
&+\Delta\|x\nu\|_{L^2}\,\mathbb{E}\left[\|m_{\Delta,h}\psi'(\varphi_{\Delta,n}-\varphi_{\Delta})\|_{L^2}\right]\\
\lesssim&n^{-1/2}\Big(2\Delta^{1/2}\|x\nu\|_{L^2}+\|x^2\nu\|_{L^2}+\Delta\|x\nu\|_{L^2}\|x\nu\|_{L^1}\Big)\|m_{\Delta,h}\|_{L^2}\\
\lesssim&n^{-1/2}h^{-\Delta\alpha-1/2}.
\end{aligned}$$

It remains to estimate the main stochastic term $M^{\nu}_{\Delta,n}$. Since U is bounded, we find a finite number of points $t_1,\dots,t_{L_n}\in U$ such that $\sup_{t\in U}\min_{l=1,\dots,L_n}|t-t_l|\leqslant(n\Delta)^{-2}$

and $L_n \lesssim (n\Delta)^2$. The inequalities by Young and by Cauchy–Schwarz yield

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in U}\min_{l=1,\dots,L_n}|M^{\nu}_{\Delta,n}(t)-M^{\nu}_{\Delta,n}(t_l)|\Big]\\
\lesssim&(n\Delta)^{-2}\Delta^{-1}\,\mathbb{E}\Big[\big\|(K_h')*\big(\mathcal{F}\left[\mathbb{1}_{[-1/h,1/h]}\varphi_\Delta^{-1}(\varphi''_{\Delta,n}-\varphi''_\Delta)\right]\big)\big\|_\infty\Big]\\
\lesssim&(n\Delta)^{-2}\Delta^{-1}h^{-1}\|K'\|_{L^1}\int_{-1/h}^{1/h}|\varphi_\Delta^{-1}|(u)\,\mathbb{E}[|\varphi''_{\Delta,n}-\varphi''_\Delta|(u)]\mathrm{d}u\\
\lesssim&(n\Delta)^{-2}h^{-\Delta\alpha-2}.
\end{aligned}
$$

Together with Markov's inequality and Lemma 5.22, defining $T:=\sup_{t\in U}|t|$ this yields for $n\Delta$ sufficiently large

$$
\begin{aligned}
&P\Big(\sup_{t\in U}|M^{\nu}_{\Delta,n}(t)|>\kappa_0\Big(\frac{\log(n\Delta)}{n\Delta}\Big)^{1/2}h^{-\Delta\alpha-1/2}\Big)\\
\leqslant&P\Big(\max_{l=1,\dots,L_n}|M^{\nu}_{\Delta,n}(t_k)|>\frac{\kappa_0}{2}\Big(\frac{\log(n\Delta)}{n\Delta}\Big)^{1/2}h^{-\Delta\alpha-1/2}\Big)\\
&\quad+\frac{2}{\kappa_0}\Big(\frac{n\Delta}{\log(n\Delta)}\Big)^{1/2}h^{\Delta\alpha+1/2}\,\mathbb{E}\Big[\sup_{t\in U}\min_{l=1,\dots,L_n}|M^{\nu}_{\Delta,n}(t)-M^{\nu}_{\Delta,n}(t_l)|\Big]\\
\leqslant&2L_n\exp\Big(-\frac{c\kappa_0^2\log(n\Delta)}{2(1+T^3)(2+\kappa_0(\log(n\Delta)/(n\Delta h))^{1/2})}\Big)\\
&\quad+o\Big((n\Delta)^{-3/2}(\log n\Delta)^{-1/2}h^{-3/2}\Big)\\
\leqslant&2\exp\big((2-\tfrac{c}{6}\kappa_0^2(1+T^3)^{-1})\log(n\Delta)\big)+o(1),
\end{aligned}
$$

which converges to zero as $n\Delta\to\infty$ if κ_0 is chosen sufficiently large. □

Proof of Theorem 5.9

We prove consistency analogously to Lemma 4.16.

Lemma 5.24. *Let $\alpha,\beta,s,\zeta,r,R>0, m>4, s'\in(-1,0]$ and let $\eta_n\downarrow 0$ with $\eta_n^{-1}\lesssim\log n$. Suppose the kernel satisfies (5.6) with order $p\geqslant 1$. Then we have*

(i) for any bandwidth satisfying $(\log n)^4h^{1+s'}\to 0$ and $(n\Delta)^{-1}h^{-2\Delta\alpha-1}\to 0$

$$\sup_{(\sigma^2,\gamma,\nu)\in\widetilde{\mathcal{D}}^{s,s'}_\tau(\alpha,m,\zeta,\eta_n,R)}P\big(|\widehat{q}^{\pm}_{\tau,h}-q^{\pm}_\tau|>\delta\big)\to 0\quad\text{for all }\delta>0,$$

(ii) for any bandwidth satisfying $(\log n)^4h^{1+s'}\to 0$ and $(\log n)^4(n\Delta)^{-1}h^{-2}e^{2r\Delta h^{-\beta}}\to 0$

$$\sup_{(\sigma^2,\gamma,\nu)\in\widetilde{\mathcal{E}}^{s,s'}_\tau(\beta,m,\zeta,\eta_n,r,R)}P\big(|\widehat{q}^{\pm}_{\tau,h}-q^{\pm}_\tau|>\delta\big)\to 0\quad\text{for all }\delta>0.$$

Proof. We adopt the general strategy of the proof of Theorem 5.7 by van der Vaart (1998) in the classical M-estimation setting. Without loss of generality, we only consider $\widehat{q}^{+}_{\tau,h}$.

Step 1: By the Hölder regularity we have $\nu(t)\geqslant\nu(q_\tau^+)-|\nu(q_\tau^+)-\nu(t)|\geqslant\frac{1}{R}-R|q_\tau^+-t|^{1\wedge\alpha}\geqslant\frac{1}{2R}$ for $|q_\tau^+-t|\leqslant(2R^2)^{-(1\vee\alpha^{-1})}$. Without loss of generality we can assume $\delta\leqslant\delta_0:=(2R^2)^{-(1\vee\alpha^{-1})}\wedge(\eta_n/2)\wedge\zeta$, otherwise consider $\delta\wedge\delta_0$. Using $N(q_\tau^+)=\tau$ and monotonicity of N, we obtain the uniqueness condition

$$\inf_{t>\eta_n:|t-q_\tau^+|\geqslant\delta}|N(t)-\tau|\geqslant\delta\inf_{t>\eta_n:|t-q_\tau^+|\leqslant\delta}\nu(t)\geqslant\frac{\delta}{2R}.$$

Step 2: Since $\widehat{q}_{\tau,h}^+$ minimizes $|\widehat{N}_h-\tau|$ on the interval (η_n,∞) and $N(q_\tau^+)=\tau$ with $q_\tau\in(\eta_n,\infty)$, Proposition 5.7 and the assumptions on h imply for any $\delta>0$

$$P\big(|\widehat{N}_h(\widehat{q}_{\tau,h}^+)-\tau|>\delta\big)\leqslant P\big(|\widehat{N}_h(q_\tau^+)-N(q_\tau^+)|>\delta\big)\to 0.$$

We could even find an event A such that $P(A)\to 0$ and $\widehat{N}_h(\widehat{q}_\tau)-\tau=0$ on A exactly as for (4.38) in the deconvolution setting.

Step 3: We infer from Steps 1 and 2

$$\begin{aligned}P\big(|\widehat{q}_{\tau,h}^+-q_\tau^+|>\delta\big)&\leqslant P\big(|N(\widehat{q}_{\tau,h}^+)-\tau|\geqslant\delta/(2R)\big)\\&=P\big(|N(\widehat{q}_{\tau,h}^+)-\widehat{N}_h(\widehat{q}_{\tau,h}^+)|\geqslant\delta/(3R)\big)+o(1)\\&\leqslant P\big(\sup_{t\in(\eta_n,\infty)}|N(t)-\widehat{N}_h(t)|\geqslant\delta/(3R)\big)+o(1).\end{aligned}\tag{5.42}$$

Hence, it remains to show uniform consistency of $\widehat{N}_h(t)$. Applying the error decomposition (5.22) $|N(t)-\widehat{N}_h(t)|\leqslant|B_n(t)|+|S_n(t)|+|V_n(t)|$ and the estimates in Lemma 5.17 and Proposition 5.18 we obtain

$$\sup_{t\in(\eta_n,\infty)}|V_n(t)|\lesssim\eta_n^{-4}h,\qquad\sup_{t\in(\eta_n,\infty)}|B_n(t)|\lesssim\eta_n^{-4}h^{1+s'}.$$

According to Lemma 5.16 and Proposition 5.19, the stochastic error term can be decomposed into $S_n(t)=M_{\Delta,n}(t)+(L_{\Delta,n}-M_{\Delta,n})(t)+R_n(t)$, where $\sup_{|t|\geqslant\eta_n}|L_{\Delta,n}-M_{\Delta,n}|(t)+|R_n|(t)$ is of the order claimed in Proposition 5.19. For $(\sigma^2,\gamma,\nu)\in\mathcal{E}^s(\beta,m,U,r,R)$ the main stochastic error term is uniformly bounded as well. For the case $(\sigma^2,\gamma,\nu)\in\mathcal{D}^s(\alpha,m,U,R)$ it remains to apply Proposition 5.21 on an appropriate grid. For $\kappa\in(0,1)$ we define a grid $v_l=-\kappa^{-1}+l\kappa$ with $l=0,\dots,L:=2\lfloor\kappa^{-2}\rfloor$. Set $t_l:=\operatorname{sign}(v_l)(|v_l|\vee\eta_n)$ for $l=0,\dots,L$. Then

$$\sup_{|t|\in(\eta_n,\infty)}|M_{\Delta,n}(t)|\leqslant\sup_{l=0,\dots,L}|M_{\Delta,n}(t_l)|+\sup_{t\in(\eta_n,\infty)}\min_{l=0,\dots,L}|M_{\Delta,n}(t)-M_{\Delta,n}(t_l)|.$$

Noting that $\|g_t-g_s\|_{L^1}\leqslant|t-s|/(s^2\wedge t^2)$ and $\|g_t-g_s\|_{L^1}\leqslant 2\int_{|t|\wedge|s|}^\infty x^{-2}\mathrm{d}x\sim|t|^{-1}\vee|s|^{-1}$, increments of $M_{\Delta,n}(t)$ can be estimated using Plancherel's identity and Fubini's theorem

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in(\eta_n,\infty)}\min_{l=0,\dots,L}|M_{\Delta,n}(t)-M_{\Delta,n}(t_l)|\Big]\\
\leqslant&\frac{1}{2\pi\Delta}\,\mathbb{E}\Big[\sup_{t\in(\eta_n,\infty)}\min_{l=0,\dots,L}\Big|\int\mathcal{F}[g_t-g_{t_l}](-u)m_{\Delta,h}(u)(\varphi''_{\Delta,n}(u)-\varphi''_{\Delta}(u))\mathrm{d}u\Big|\Big]\\
\leqslant&\frac{1}{2\pi\Delta}\sup_{t\in(\eta_n,\infty)}\min_{l=0,\dots,L}\|g_t-g_{t_l}\|_{L^1}\int_{-1/h}^{1/h}|m_{\Delta,h}(u)|\,\mathbb{E}[(\varphi''_{\Delta,n}(u)-\varphi''_{\Delta}(u))^2]^{1/2}\mathrm{d}u\\
\lesssim&(n\Delta)^{-1/2}\sup_{t\in(\eta_n,\infty)}\min_{l=0,\dots,L}\|g_t-g_{t_l}\|_{L^1}\|(1+|u|)^{\Delta\alpha}\|_{L^1(I_h)}\\
\lesssim&(n\Delta)^{-1/2}\eta_n^{-2}\kappa h^{-\Delta\alpha-1}.
\end{aligned}$$

Choosing $\kappa=\eta_n^2(n\Delta)^{-1/2}$, Markov's inequality and Proposition 5.21 yield for Δn sufficiently large and some constant $c>0$

$$\begin{aligned}
&P\Big(\sup_{|t|\in(\eta_n,\infty)}|M_{\Delta,n}(t)|>\delta\Big)\\
\leqslant&P\Big(\sup_{l=0,\dots,L}|M_{\Delta,n}(t_j)|>\frac{\delta}{2}\Big)+\frac{2}{\delta}\,\mathbb{E}\Big[\sup_{t\in(\eta_n,\infty)}\min_{l=0,\dots,L}|M_{\Delta,n}(t)-M_{\Delta,n}(t_l)|\Big]\\
\leqslant&2(L+1)\exp\Big(-c\delta^2\eta_n^3 n\Delta\|(1+|u|)^{\Delta\alpha-1}\|_{L^2(I_h)}^{-2}\Big)+(n\Delta)^{-1/2}\eta_n^{-2}\kappa h^{-\Delta\alpha-1}\\
\leqslant&2\eta_n^{-4}n\Delta\exp(-c\delta^2\eta_n^3 n\Delta h^{2\Delta\alpha})+(n\Delta)^{-1}h^{-\Delta\alpha-1}\to0,
\end{aligned}$$

owing to $n\Delta h^{2\Delta\alpha+1}\to\infty$. □

Proof of Theorem 5.9. Proposition 5.7 shows that the numerator in the error representation (5.8) is of the claimed order. Moreover, it holds for any $\delta>0$

$$P(|\widehat{\nu}_h(\xi^{\pm})-\nu(q_\tau^{\pm})|>\delta)\leqslant P(\sup_{|t|<\zeta}|\widehat{\nu}_h(q_\tau^{\pm}+t)-\nu(q_\tau^{\pm})|>\delta)+P(|\widehat{q}_{\tau,h}^{\pm}-q_\tau^{\pm}|\geqslant\zeta),$$

where the first term converges to zero by Proposition 5.23 and the second one tends to zero by Lemma 5.24. Therefore, the denominator in (5.8) can be written as

$$\widehat{\nu}_h(\xi^{\pm})=\nu(q_\tau^{\pm})+o_P(1).$$ □

5.4.2. Proofs for Section 5.3

To prove Lemma 5.11, we apply some entropy arguments. To fix the notation, we define for any (pseudo-)metric d on $\mathbb{R}$ the covering number $N(r,A,d)$ as the smallest number of balls with radius $r>0$ which is necessary to cover a subset $A\subseteq\mathbb{R}$. For $v>0$ the entropy integral is defined as

$$J(v,A,d):=\int_0^v\sqrt{\log N(r,A,d)}\mathrm{d}r,$$

which is finite for any v if $N(r,A,d)$ grows polynomially in r^{-1}.

Proof of Lemma 5.11. Itô's isometry yields

$$\begin{aligned}\operatorname{Var}(\Phi_n(u)) = \mathbb{E}[|\Phi_n(u)|^2] &= n^{-1}u^2(u^2+1)\,\mathbb{E}\Big[\Big|\int e^{iux-x}\rho(x)\mathrm{d}W(x)\Big|^2\Big]\\ &= n^{-1}u^2(u^2+1)\|e^{-x}\rho(x)\|_{L^2}^2 \qquad (5.43)\end{aligned}$$

and similarly for $\Phi_n^{(1)}$ and $\Phi_n^{(2)}$ as defined in (5.16). From Itô's isometry and dominated convergence we conclude that $\Phi_n^{(k)}, k = 1,2$, are the first and second order $L^2(P)$-derivatives of Φ_n. The intrinsic covariance metric of the Gaussian process

$$\Big\{\int e^{iux-x}x^k\rho(x)\mathrm{d}W(x) : u \in \mathbb{R}\Big\}$$

is given by

$$d^{(k)}(u,v) := \mathbb{E}\Big[\Big|\int (e^{iux}-e^{ivx})x^k e^{-x}\rho(x)\mathrm{d}W(x)\Big|^2\Big]^{1/2}.$$

Using $\int |x|^m e^{-2x}\rho^2(x)\mathrm{d}x < \infty$, the entropy integrals $J(\infty, [-U,U], d^{(k)})$ can be bounded exactly as in the proof of Proposition 1 by Söhl (2014) and are of order $\sqrt{\log U}$. Dudley's theorem (e.g. Massart, 2007, Prop. 3.18) yields then $\mathbb{E}[\|\Phi_n^{(k)}\|_{L^\infty[-U,U]}] \lesssim n^{-1/2}U^2\sqrt{\log U})$. □

Convergence rates

Since the proof strategy is the same for the observation schemes in Sections 5.2 and 5.3, we will concentrate on the differences. The estimation error can be decomposed into $\widetilde{N}_t(t) - N(t) = B_n(t) + \widetilde{S}_n(t) + V_n(t)$ where $B_n(t)$ and $V_n(t)$ are given in (5.22) and only the stochastic error

$$\widetilde{S}_n(t) = \int g_t(x)\,\mathcal{F}^{-1}\Big[\Big(\psi''(u) - \widetilde{\psi}_n''(u)\Big)\,\mathcal{F}K(hu)\Big](x)\mathrm{d}x$$

has a different probabilistic structure. We can apply Lemma 5.17 and Proposition 5.18 to estimate $B_n(t)$ and $V_n(t)$. For the sake of brevity, we will frequently write

$$(\varphi_T^{-1}(\widetilde{\varphi}_{T,n} - \varphi_T))'' = \varphi_T^{-1}\Phi_n^{(2)} + 2(\varphi_T^{-1})'\Phi_n^{(1)} + (\varphi_T^{-1})''\Phi_n.$$

This equality is justified in $L^2(P)$-sense, but should merely be understood as notational convention. Linearizing the stochastic error term, we define analogously to (5.25)

$$\widetilde{L}_{T,n}(t) := -\frac{1}{T}\int g_t(x)\mathcal{F}^{-1}[\mathcal{F}K(h\bullet)(\varphi_T^{-1}(\widetilde{\varphi}_{T,n} - \varphi_T))''](x)\mathrm{d}x.$$

Since the $\Phi_n^{(k)}$ are almost surely bounded and $\mathcal{F}K$ has compact support, $\widetilde{L}_{T,n}$ is almost surely well defined. The remainder $\widetilde{S}_n - \widetilde{L}_{T,n}$ will be bounded on the event

$$\Omega_{n,h} := \Big\{\inf_{u\in I_h}|\widetilde{\varphi}_{T,n}(u)| \geqslant n^{-1/2}h^{-2}(\log h^{-1})\Big\}. \qquad (5.44)$$

The order of the remainder in the next lemma corresponds exactly to Lemma 5.16 taking the bound from Lemma 5.11 into account.

Lemma 5.25. *If $\int |x|^m e^{-2x}\rho^2(x)\mathrm{d}x < \infty$ for some $m > 4$, then for any sequence $h = h_n$ satisfying $n^{-1/2}h^{-2}(\log h^{-1})\|\varphi_T^{-1}\|_{L^\infty(I_h)} \to 0$ as $n \to \infty$ it holds $P(\Omega_{n,h}) \to 1$ and uniformly for all Lévy triplets (σ^2, γ, ν)*

$$\begin{aligned}&\sup_{u\in I_h}\Big|\widetilde{\psi}_n''(u) - \psi''(u) - T^{-1}(\varphi_T^{-1}(\widetilde{\varphi}_{T,n} - \varphi_T))''(u)\Big|\\&= \mathcal{O}_P\Big((1+\|\psi'\|^2_{L^\infty(I_h)})n^{-1}h^{-4}\log(h^{-1})\|\varphi_T^{-1}\|^2_{L^\infty(I_h)}\Big).\end{aligned}$$

Proof. $P(\Omega_{n,h}) \to 1$ can be verified exactly as in Lemma 4.13. To bound $\widetilde{\varphi}_{T,n}^{-1} - \varphi_T^{-1}$, we apply the argument by Neumann (1997), cf. also (4.23), and Lemma 5.11. We obtain on $\Omega_{n,h}$

$$\begin{aligned}\|\widetilde{\varphi}_{T,n}^{-1} - \varphi_T^{-1}\|_{L^\infty(I_h)}\mathbb{1}_{\Omega_{n,h}} &\leqslant \sup_{u\in I_h}\Big(\frac{|\Phi_n(u)|}{|\varphi_T(u)|^2} + n^{1/2}h^2|\log h|^{-1}\frac{|\Phi_n(u)|^2}{|\varphi_T(u)|^2}\Big)\\&= \mathcal{O}_P\big(n^{-1/2}h^{-2}|\log h|^{1/2}\|\varphi_T^{-1}\|^2_{L^\infty(I_h)}\big).\end{aligned} \tag{5.45}$$

A straight forward computation shows on $\Omega_{n,h}$

$$\begin{aligned}T(\widetilde{\psi}_n'' - \psi'') &= \frac{\widetilde{\varphi}_{T,n}''}{\widetilde{\varphi}_{T,n}} - \frac{\varphi_T''}{\varphi_T} - \Big(\frac{\widetilde{\varphi}_{T,n}'}{\widetilde{\varphi}_{T,n}}\Big)^2 + \Big(\frac{\varphi_T'}{\varphi_T}\Big)^2\\
=&\frac{\widetilde{\varphi}_{T,n}'' - \varphi_T''}{\varphi_T} - \Big(\frac{\widetilde{\varphi}_{T,n}'}{\widetilde{\varphi}_{T,n}} + \frac{\varphi_T'}{\varphi_T}\Big)\frac{\widetilde{\varphi}_{T,n}' - \varphi_T'}{\varphi_T} + \Big(\frac{\widetilde{\varphi}_{T,n}''}{\widetilde{\varphi}_{T,n}} - \Big(\frac{\widetilde{\varphi}_{T,n}'}{\widetilde{\varphi}_{T,n}} + \frac{\varphi_T'}{\varphi_T}\Big)\frac{\widetilde{\varphi}_{T,n}'}{\widetilde{\varphi}_{T,n}}\Big)\frac{\varphi_T - \widetilde{\varphi}_{T,n}}{\varphi_T}\\
=&\frac{\widetilde{\varphi}_{T,n}'' - \varphi_T''}{\varphi_T} - 2\Big(\frac{\varphi_T'}{\varphi_T}\Big)\frac{\widetilde{\varphi}_{T,n}' - \varphi_T'}{\varphi_T} + \Big(\frac{\varphi_T''}{\varphi_T} - 2\frac{\varphi_T'}{\varphi_T}\frac{\varphi_T'}{\varphi_T}\Big)\frac{\varphi_T - \widetilde{\varphi}_{T,n}}{\varphi_T} + R_n\\
=&\varphi_T^{-1}\Phi_n^{(2)} + 2(\varphi_T^{-1})'\Phi_n^{(1)} + (\varphi_T^{-1})''\Phi_n + R_n\end{aligned} \tag{5.46}$$

where R_n is the sum of all second order terms. Using (5.45), formulas (5.19) as well as Lemma 5.11, we obtain the claimed order of R_n. □

Let us study the linearized stochastic error term. To apply the Lepski method later we need a sharp bound on the variance of $\widetilde{L}_{T,n}$. For the asymptotic analysis we need the order of magnitude. With the auxiliary functions, $u \in \mathbb{R}$,

$$\begin{aligned}\chi_t^{(0)}(u) &:= \mathcal{F}g_t(-u)\,\mathcal{F}K(hu)\big(u(u-i)(\varphi_T^{-1})''(u) + 2(2u-i)(\varphi_T^{-1})'(u) + 2\varphi_T^{-1}(u)\big),\\
\chi_t^{(1)}(u) &:= \mathcal{F}g_t(-u)\,\mathcal{F}K(hu)\big(2u(iu+1)(\varphi_T^{-1})'(u) + (4iu+2)\varphi_T^{-1}(u)\big),\\
\chi_t^{(2)}(u) &:= u(i-u)\,\mathcal{F}g_t(-u)\,\mathcal{F}K(hu)\varphi_T^{-1}(u)\end{aligned}$$

we define

$$\begin{aligned}\Sigma_{n,h}(t) :=&\frac{1}{2\pi n^{1/2}T}\Big(\|x^2e^{-x}\rho(x)\|_\infty\|\chi_t^{(2)}\|_{L^2}\\&\qquad + \|xe^{-x}\rho(x)\|_\infty\|\chi_t^{(1)}\|_{L^2} + \|e^{-x}\rho(x)\|_\infty\|\chi_t^{(0)}\|_{L^2}\Big).\end{aligned} \tag{5.47}$$

Lemma 5.26. *If $\int(1+|x|)^4e^{-2x}\rho^2(x)\mathrm{d}x < \infty$, then $\widetilde{L}_{T,h}(t)$ is centered normal. Supposing additionally $\|(1\vee x^2)e^{-x}\rho\|_\infty \lesssim 1$, it holds*

$$\mathbb{E}[|\widetilde{L}_{T,n}(t)|^2]^{1/2} \leqslant \Sigma_{n,h}(t) \lesssim n^{-1/2}(t^{-1}\vee t^{-2})\big\|(1+|u|)|\varphi_T(u)|^{-1}(1+\psi'(u)^2)\big\|_{L^2(I_h)}.$$

Proof. Since the $\Phi_n^{(k)}$ are almost surely bounded, we can apply Plancherel's identity which yields

$$\begin{aligned}\mathbb{E}[|\widetilde{L}_{T,n}(t)|^2] = \frac{1}{(2\pi T)^2}\,\mathbb{E}\Big[\Big|\int \mathcal{F}\,g_t(-u)\,\mathcal{F}\,K(hu)\varphi_T^{-1}(u)\Phi_n^{(2)}(u)\mathrm{d}u\\ + 2\int \mathcal{F}\,g_t(-u)\,\mathcal{F}\,K(hu)(\varphi_T^{-1})'(u)\Phi_n^{(1)}(u)\mathrm{d}u\\ + \int \mathcal{F}\,g_t(-u)\,\mathcal{F}\,K(hu)(\varphi_T^{-1})''(u)\Phi_n(u)\mathrm{d}u\Big|^2\Big].\end{aligned}$$

By continuity and boundedness of $\Phi_n^{(k)}, k = 0, 1, 2$, the integral in u can be approximated with Riemann sums. We conclude first that $\widetilde{L}_{T,h}(t)$ is normally distributed, cf. Section 6.2 in Söhl (2014), and second that we can exchange the deterministic integral and the stochastic integral due to the construction of the Wiener integral as $L^2(P)$-limit. Together with Itô's isometry and Plancherel's identity we obtain

$$\begin{aligned}&\mathbb{E}[|\widetilde{L}_{T,n}(t)|^2]^{1/2}\\ =&\frac{1}{n^{1/2}T}\,\mathbb{E}\Big[\Big|\int (x^2\,\mathcal{F}^{-1}\,\chi_t^{(2)}(-x) + x\mathcal{F}^{-1}\chi_t^{(1)}(-x) + \mathcal{F}^{-1}\,\chi_t^{(0)}(-x))e^{-x}\rho(x)\mathrm{d}W(x)\Big|^2\Big]^{1/2}\\ =&\frac{1}{n^{1/2}T}\Big(\int \Big|x^2\,\mathcal{F}^{-1}\,\chi_t^{(2)}(-x) + x\mathcal{F}^{-1}\chi_t^{(1)}(-x) + \mathcal{F}^{-1}\,\chi_t^{(0)}(-x)\Big|^2 e^{-2x}\rho^2(x)\mathrm{d}x\Big)^{1/2}\\ \leqslant&\frac{1}{2\pi n^{1/2}T}\Big(\|x^2e^{-x}\rho(x)\|_\infty\|\chi_t^{(2)}\|_{L^2} + \|xe^{-x}\rho(x)\|_\infty\|\chi_t^{(1)}\|_{L^2} + \|e^{-x}\rho(x)\|_\infty\|\chi_t^{(0)}\|_{L^2}\Big).\end{aligned}$$

The formulas (5.19) and $|\,\mathcal{F}\,g_t(u)| \lesssim (t^{-1}\vee t^{-2})(1+|u|)^{-1}$ yield the claimed asymptotic bound. □

Now, we can conclude convergence rates for the distribution function estimator $\widetilde{N}_h$.

Proposition 5.27. *Suppose $\|(1\vee x^2)e^{-x}\rho\|_\infty \lesssim 1$ and $\int(1+|x|)^m e^{-2x}\rho^2(x)\mathrm{d}x < \infty$ for some $m > 4$. Let $U \subseteq \mathbb{R}$ be an open set and $\alpha, \beta, s, r, R > 0$. Let the kernel satisfy* (5.6) *with order $p \geqslant s+1$.*

(i) If $(\sigma^2,\gamma,\nu) \in \mathcal{D}^s(\alpha, 2, U, R)$, then $|\widetilde{N}_h(t) - N(t)| = \mathcal{O}_P\big(n^{-(s+1)/(2s+2\Delta\alpha+5)}\big)$ for $h = h_n = n^{-1/(2s+2\Delta\alpha+5)}$.

(ii) If $(\sigma^2,\gamma,\nu) \in \mathcal{E}^s(\beta, 2, U, r, R)$, then $|\widetilde{N}_h(t) - N(t)| = \mathcal{O}_P\big((\log n)^{-(s+1)/\beta}\big)$ for $h = h_n = (\frac{1}{2r}\log n)^{-1/\beta}$.

Proof. As in the proof of Proposition 5.7 we have

$$|\widetilde{N}_h(t) - N(t)| \leqslant |B_n(t)| + |\widetilde{S}_n(t)| + |V_n(t)| \lesssim h^{s+1} + |\widetilde{S}_n(t)|. \tag{5.48}$$

Lemmas 5.25 and 5.26 yield for the stochastic error term

$$\begin{aligned}|\widetilde{S}_n(t)| &\leqslant |\widetilde{L}_n(t)| + |\widetilde{S}_n(t) - \widetilde{L}_n(t)|\\ &= \mathcal{O}_P\Big(n^{-1/2}(1 + \|\psi'\|^2_{L^\infty(I_h)})\\ &\qquad \times\Big(\|(1+|u|)\varphi_T(u)|^{-1}\|_{L^2(I_h)} + n^{-1/2}h^{-4}(\log h^{-1})\|\varphi_\Delta^{-1}\|^2_{L^\infty(I_h)}\Big)\Big).\end{aligned}$$

In situation (i) we use Proposition 5.1 to see that $x\nu$ is integrable and thus

$$
\begin{aligned}
|\widetilde{N}_h - N|(t) &= \mathcal{O}_P\Big(h^{s+1} + n^{-1/2}(\|(1+|u|)^{\Delta\alpha+1}\|_{L^2(I_h)} + n^{-1/2}(\log h^{-1})h^{-2\Delta\alpha-4})\Big)\\
&= \mathcal{O}_P\Big(h^{s+1} + n^{-1/2}(h^{-\Delta\alpha-3/2} + n^{-1/2}(\log h^{-1})h^{-2\Delta\alpha-4})\Big). \qquad (5.49)
\end{aligned}
$$

Therefore, the optimal bandwidth $h_n = n^{-1/(2s+2\Delta\alpha+5)}$ yields the claimed rate.

For exponentially decaying characteristic functions in case (ii) we use $\|\psi'\|_{L^\infty(I_h)} \lesssim h^{-1}$ to infer

$$
\begin{aligned}
|\widetilde{N}_h - N|(t) =&\mathcal{O}_P\Big(h^{s+1} + n^{-1/2}(h^{-2}\|(|u|+1)e^{r|u|^\beta}\|_{L^2(I_h)} + n^{-1/2}h^{-6}(\log h^{-1})e^{rh^{-\beta}})\Big)\\
=&\mathcal{O}_P\Big(h^{s+1} + n^{-1/2}h^{-7/2}(1+n^{-1/2}h^{-5/2}(\log h^{-1}))e^{rh^{-\beta}}\Big), \qquad (5.50)
\end{aligned}
$$

leading to the rate optimal choice $h_n = (\frac{1}{2r}\log n)^{-1/\beta}$. □

Proposition 5.12 on the density estimator $\widetilde{\nu}_h$ is an immediate consequence from the following proposition.

Proposition 5.28. *Suppose $\|(1\vee x^2)e^{-x}\rho\|_\infty \lesssim 1$ and $\int(1+|x|)^m e^{-2x}\rho^2(x)\mathrm{d}x < \infty$ for some $m > 4$. Let the kernel satisfy (5.6) with order $p \geqslant s$, let $U \subseteq \mathbb{R}$ be a bounded, open set which is bounded away from zero and let $\alpha, \beta, r, R > 0$. Then we have*

(i) for any $h \downarrow 0$ satisfying $n^{-1/2}h^{-\Delta\alpha-5/2-\delta} \to 0$ for some $\delta > 0$ we have uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{D}^s(\alpha, 2, U, R)$

$$\sup_{t\in U}|\widetilde{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{D}^s}\Big(h^s + n^{-1/2}(\log n)^{1/2}h^{-\Delta\alpha-5/2}\Big),$$

(ii) for any $h \downarrow 0$ satisfying $n^{-1/2}e^{(r+\delta)h^{-\beta}} \to 0$ for some $\delta > 0$ we have uniformly in $(\sigma^2, \gamma, \nu) \in \mathcal{E}^s(\beta, 2, U, r, R)$

$$\sup_{t\in U}|\widetilde{\nu}_h(t) - \nu(t)| = \mathcal{O}_{P,\mathcal{E}^s}\Big(h^s + n^{-1/2}(\log n)^{1/2}h^{-9/2}e^{rh^{-\beta}}\Big).$$

Proof. As in the proof of Proposition 5.23 we deduce from Lemma 5.25 that

$$
\begin{aligned}
\sup_{t\in U}|\widetilde{\nu}_h(t) - \nu(t)| =&\mathcal{O}(h^s) + \sup_{t\in U}|\widetilde{L}_{T,n,\nu}(t)|\\
&+ \mathcal{O}_P\Big((1+\|\psi'\|^2_{L^\infty(I_h)})n^{-1}h^{-5}\log(h^{-1})\|\varphi_\Delta^{-1}\|^2_{L^\infty(I_h)}\Big)
\end{aligned}
$$

with linearized stochastic error term

$$\widetilde{L}_{T,n,\nu}(t) := -\frac{1}{Tt^2}\,\mathcal{F}^{-1}\Big[\mathcal{F}\,K(h\bullet)\Big(\frac{\widetilde{\varphi}_{T,n} - \varphi_T}{\varphi_T}\Big)''\Big](t).$$

Because U is bounded away from zero, it suffices to bound

$$\begin{aligned}\mathbb{E}\big[\sup_{t\in U}|t^2\widetilde{L}_{T,n,\nu}(t)|\Big] \leqslant T^{-1}\,\mathbb{E}\Big[\sup_{t\in U}|\mathcal{F}^{-1}[m_{T,h}\Phi_n^{(2)}](t)|] + \mathbb{E}[\sup_{t\in U}|\mathcal{F}^{-1}[m_{T,h}\psi'\Phi_n^{(1)}](t)|]\\ + \mathbb{E}[\sup_{t\in U}|\mathcal{F}^{-1}[m_{T,h}(\psi''-T(\psi')^2)\Phi_n](t)|\Big]\\ =: E_1+E_2+E_3.\end{aligned}$$

Since all three terms can be estimated analogously, we limit ourselves on E_3 which has the largest variance. We estimate uniformly in $t\in U$ with use of Plancherel's identity, Fubini's theorem and Itô's isometry

$$\begin{aligned}&\operatorname{Var}\big(\mathcal{F}^{-1}[m_{T,h}(\psi''-T(\psi')^2)\Phi_n](t)\big)\\ =&\frac{1}{2\pi}\,\mathbb{E}\Big[\Big|\int e^{-itu}m_{T,h}(u)(\psi''(u)-T\psi'(u)^2)\Phi_n(u)\mathrm{d}u\Big|^2\Big]\\ =&n^{-1}\,\mathbb{E}\Big[\Big|\int\mathcal{F}^{-1}[m_{T,h}(u)(\psi''(u)-T\psi'(u)^2)u(u-i)e^{-itu}](-x)\rho(x)\mathrm{d}W(x)\Big|^2\Big]\\ =&n^{-1}\int\big|\,\mathcal{F}^{-1}[m_{T,h}(u)(\psi''(u)-T\psi'(u)^2)u(u-i)e^{-itu}](-x)\rho(x)\big|^2\mathrm{d}x\\ \lesssim&n^{-1}\int(u^2+u^4)|m_{T,h}(u)|^2(1+|\psi'(u)|^2)^2\mathrm{d}u =: v(n,h).\end{aligned}\tag{5.51}$$

Analogously, we can estimate the distance in the intrinsic norm for any $\delta\in(0,1/2)$

$$\begin{aligned}d(s,t)^2 :=& \mathbb{E}\Big[\Big|\mathcal{F}^{-1}[m_{T,h}(\psi''-T(\psi')^2)\Phi_n](t)-\mathcal{F}^{-1}[m_{T,h}(\psi''-T(\psi')^2)\Phi_n](s)\Big|^2\Big]\\ =&n^{-1}\int(1+u^4)|m_{T,h}(u)|^2(1+|\psi'(u)|^2)^2|e^{-itu}-e^{-isu}|^2\mathrm{d}u\\ \lesssim&|t-s|^{2\delta}n^{-1}\int(1+u^{4+2\delta})|m_{T,h}(u)|^2(1+|\psi'(u)|^2)^2\mathrm{d}u\\ =:&|t-s|^{2\delta}c_\delta(n,h)\end{aligned}$$

and thus the covering number is of the order $N(r,U,d)\lesssim(c_\delta(n,h)/r)^{1/(2\delta)}$. Consequently, the entropy integral can be bounded by

$$\begin{aligned}J(\sqrt{v(n,h)},U,d) &\lesssim \int_0^{\sqrt{v(n,h)}}\big(\log c_\delta(n,h)+\log r^{-1}\big)^{1/2}\mathrm{d}r\\ &\lesssim v(n,h)^{1/2}\big(\log c_\delta(n,h)+\log v(n,h)^{-1}\big)^{1/2}.\end{aligned}\tag{5.52}$$

Using this entropy bound, Dudley's theorem (e.g. Massart, 2007, Prop. 3.18) yields

$$\mathbb{E}\Big[\sup_{t\in U}|\,\mathcal{F}^{-1}[m_{T,h}(\psi''-T(\psi')^2)\Phi_n](t)|\Big]\lesssim v(n,h)^{1/2}\big(\log c_\delta(n,h)+\log v(n,h)^{-1}\big)^{1/2}.$$

Now we can plug in the different assumptions on the decay of φ_T (and in particular $\log c_\delta(n,h)$ is smaller than 0 for δ sufficiently small). $\square$

To finally prove Theorem 5.13, we can argue as for Theorem 5.9. The only ingredient which remains be shown is uniform convergence $\sup_{|t|>\eta}|\widetilde{N}_h(t)-N(t)|=o_P(1)$ for the rate optimal bandwidth $h=h_n$. Since the remainder $|\widetilde{S}_n(t)-\widetilde{L}_n(t)|$ can be bounded uniformly in t using Lemma 5.25, it suffices to show:

Lemma 5.29. *If $\|(1 \vee x^2)e^{-x}\rho\|_\infty \lesssim 1$ and $\int(1+|x|)^m e^{-2x}\rho^2(x)\mathrm{d}x < \infty$ for some $m > 4$, then it holds uniformly for all (σ^2, γ, ν)*

$$\mathbb{E}[\sup_{|t|\geqslant\eta} |\widetilde{L}_{T,n}(t)|] \lesssim \eta^{-2} n^{-1/2} (\log n)^{1/2} \Big(\int |m_{T,h}(u)|^2 (1+|\psi'(u)|^2)(1+u^4)\mathrm{d}u \Big)^{1/2}.$$

Proof. Let us estimate the covering number of $\mathbb{R}\setminus(-\eta,\eta)$ with respect to the intrinsic metric of the process $\widetilde{L}_{T,n}(t)$. Similarly to Proposition 5.28 we infer

$$\begin{aligned} d(s,t) :&= \mathbb{E}[|\widetilde{L}_{T,n}(t) - \widetilde{L}_{T,n}(s)|^2]^{1/2} \\ &\lesssim n^{-1/2} \Big(\int |\mathcal{F}[g_t - g_s](-u) m_{T,h}(u)|^2 (1+|\psi'(u)|^2)(1+u^4)\mathrm{d}u \Big)^{1/2} \\ &\lesssim \|g_t - g_s\|_{L^1} n^{-1/2} \Big(\int |m_{T,h}(u)|^2 (1+|\psi'(u)|^2)(1+u^4)\mathrm{d}u \Big)^{1/2} \\ &=: \|g_t - g_s\|_{L^1} c(n,h). \end{aligned}$$

As in Lemma 5.24, we see that the covering numbers are polynomial:

$$N(r, \mathbb{R}\setminus(-\eta,\eta), d) \lesssim \eta_n^{-4} c(n,h)^{-2} r^{-2}.$$

With the variance bound $\Sigma_{n,h} := \sup_{|t|>\eta} \Sigma_{n,h}(t)$ from (5.47) Dudley's theorem and Lemma 5.26 yield

$$\begin{aligned} \mathbb{E}[\sup_{|t|\geqslant\eta} |\widetilde{L}_{T,n}(t)|] &\lesssim J(\Sigma_{n,h}, \mathbb{R}\setminus(-\eta,\eta), d) \\ &\lesssim \Sigma_{n,h} \big(\log(\eta_n^{-2} c(n^{-1/2}, h)) + \log \Sigma_{n,h}^{-1} \big)^{1/2}. \end{aligned}$$

□

Adaptive method

The proof of Theorem 5.14 is very similar to the one of Theorem 4.12 and thus we omit the details. Due to $P(\Omega_{n,h}) \to 1$ for the minimal bandwidth and with $\Omega_{n,h}$ from (5.44), it suffices to bound all terms on the complement $\Omega_{n,h}^c$. For the bandwidth set $\mathcal{B}_n$ we can verify an analogous result as Lemma 4.11. Using (5.8), an analogue to (4.38) and (5.48), the estimation error of $\widetilde{q}_{\tau,h}$ can be bounded by

$$|\widetilde{q}_{\tau,h}^{\pm} - q_\tau^{\pm}| \leqslant \frac{|B_{n,h}(q_\tau^{\pm})| + |\widetilde{S}_{n,h}(q_\tau^{\pm})| + |V_{n,h}(q_\tau^{\pm})|}{|\widetilde{\nu}_{n,h}(\xi^{\pm})|} \leqslant \frac{Dh^{s+1} + |\widetilde{S}_{n,h}(q_\tau^{\pm})|}{|\widetilde{\nu}_{n,h}(\xi^{\pm})|} \tag{5.53}$$

with probability converging to one and with a deterministic constant $D > 0$. We can verify with use of Proposition 5.28, (5.42) and Lemmas 5.25 and 5.29 for $\eta \in (0,1)$,

$$P\Big(\max_{h\in\mathcal{B}_n} \sup_{\xi^{\pm}\in[q_\tau^{\pm}\wedge\widetilde{q}_{\tau,h}^{\pm}, q_\tau^{\pm}\vee\widetilde{q}_{\tau,h}^{\pm}]} |\widetilde{\nu}_{n,h}(\xi^{\pm}) - \nu(q_\tau^{\pm})| > \eta\nu(q_\tau^{\pm}) \Big) \to 0. \tag{5.54}$$

To conclude that $\widetilde{V}_n^{\pm}(h)$ is an appropriate upper bound for $|\widetilde{S}_{n,h}(t)|/|\widetilde{\nu}_{n,h}(\xi^{\pm})|$ in (5.53), we have to control $\widetilde{S}_{n,h}(q_\tau^{\pm})$. We again decompose it into linearization $\widetilde{L}_{n,h}(q_\tau^{\pm})$ and remainder. Since $\widetilde{L}_{n,h}(q_\tau^{\pm})$ is centered and normally distributed with variance bounded by $\Sigma^2_{n,h}(q_\tau^\pm)$ from (5.47), the Gaussian concentration and $|\mathcal{B}_n| \lesssim \log n$ yield for any $\delta > 0$

$$P\Big(\exists h \in \mathcal{B}_n : |\widetilde{L}_{n,h}(q_\tau^\pm)| > (1+\delta)\sqrt{2\log\log n}\Sigma_{n,h}(q_\tau^\pm)\Big) \to 0.$$

For the remainder, one can show, using (5.46) pointwise, $\mathbb{E}[|\widetilde{S}_{n,h}(q_\tau^\pm) - \widetilde{L}_{n,t}(q_\tau^\pm)|\mathbb{1}_{\Omega_{n,h}}] \lesssim \Sigma_{n,h}(q_\tau^\pm)^2 h^{-1}$ and thus we conclude for any $\delta > 0$

$$P\Big(\exists h \in \mathcal{B}_n : |\widetilde{S}_{n,h}(q_\tau^\pm)| > (1+\delta)\sqrt{2\log\log n}\Sigma_{n,h}(q_\tau^\pm)\Big) \to 0,$$

provided that $(\log n)\Sigma_{n,h}(q_\tau^\pm)h^{-1} \to 0$. Noting that the order of $\Sigma_{n,h}(q_\tau^\pm)h^{-1}$ is the same as the order of the stochastic error of $\widetilde{\nu}_{n,h}$, this condition is satisfied for all $h \in \mathcal{B}_n$ by construction. In the next step, we show that $\Sigma_{n,h}(q_\tau^\pm)$ is reasonably estimated by $\widetilde{\Sigma}^\pm_{n,h}$ from (5.17).

Lemma 5.30. *In the situation of Theorem 5.14 we have for any sequence $h \downarrow 0$ satisfying $\inf_{|u|\leqslant 1/h}|\varphi_T(h)| > n^{-1/2}h^{-2}\log h^{-1}$ that*

$$|\widetilde{\Sigma}^\pm_{n,h} - \Sigma_{n,h}(q_\tau^\pm)| = \mathcal{O}_P\big((\log h^{-1})^{-1}\Sigma_{n,h}(q_\tau^\pm)\big).$$

Proof. Without loss of generality we only consider q_τ^+. The triangle inequality yields

$$\begin{aligned}|\widetilde{\Sigma}^+_{n,h} - \Sigma_{n,h}(q_\tau^+)| \leqslant &\frac{1}{2\pi n^{1/2}T}\Big(\|x^2e^{-x}\rho(x)\|_\infty\|\widetilde{\chi}^{(2)}_{\widetilde{q}^+_{\tau,h}} - \chi^{(2)}_{q_\tau^+}\|_{L^2}\\ &+ \|xe^{-x}\rho(x)\|_\infty\|\widetilde{\chi}^{(1)}_{\widetilde{q}^+_{\tau,h}} - \chi^{(1)}_{q_\tau^+}\|_{L^2} + \|e^{-x}\rho(x)\|_\infty\|\widetilde{\chi}^{(0)}_{\widetilde{q}^+_{\tau,h}} - \chi^{(0)}_{q_\tau^+}\|_{L^2}\Big).\end{aligned}$$

Hence, it suffices to show

$$\|\widetilde{\chi}^{(k)}_{\widetilde{q}^+_{\tau,h}} - \chi^{(k)}_{q_\tau^+}\|_{L^2} = \mathcal{O}_P\big((\log h^{-1})^{-1}\|\chi^{(k)}_{q_\tau^+}\|_{L^2}\big), \quad \text{for } k = 0,1,2. \tag{5.55}$$

Let us start with $k = 2$ where we have

$$\begin{aligned}&\int|\widetilde{\chi}^{(2)}_{\widetilde{q}^+_{\tau,h}}(u) - \chi^{(2)}_{q_\tau^+}(u)|^2\mathrm{d}u\\ =&\int(u^2+u^4)|\mathcal{F}K(hu)|^2|\mathcal{F}g_{\widetilde{q}^+_{\tau,h}}(-u)\widetilde{\varphi}^{-1}_{T,n}(u) - \mathcal{F}g_{q_\tau^+}(-u)\varphi_T^{-1}(u)|^2\mathrm{d}u\\ \leqslant&\int(u^2+u^4)|\mathcal{F}K(hu)|^2|\varphi_T(u)|^{-2}|\mathcal{F}g_{\widetilde{q}^+_{\tau,h}}(-u) - \mathcal{F}g_{q_\tau^+}(-u)|^2\mathrm{d}u\\ &+\int(u^2+u^4)|\mathcal{F}K(hu)|^2|\mathcal{F}g_{\widetilde{q}^+_{\tau,h}}(-u)|^2|\widetilde{\varphi}^{-1}_{T,n}(u) - \varphi_T^{-1}(u)|^2\mathrm{d}u\\ =:&T_1 + T_2.\end{aligned}$$

Using $(g_t - g_s)(x) = x^2\mathbb{1}_{(s,t]}$ for $0 < s < t$ and $|\mathcal{F}g_t(u)| \sim (1+|u|)^{-1}$, the first integral can be estimated by

$$\begin{aligned}T_1 &\leqslant \|g_{\widetilde{q}^+_{\tau,h}} - g_{q_\tau^+}\|^2_{L^1}\int(u^2+u^4)|\mathcal{F}K(hu)|^2|\varphi_T(u)|^{-2}\mathrm{d}u\\ &\leqslant \eta_n^{-4}|\widetilde{q}^+_{\tau,h} - q_\tau^+|^2h^{-2}\|\chi^{(2)}_{q_\tau^+}\|^2_{L^2}.\end{aligned}$$

Applying (5.49), (5.50) and (5.54), we conclude $T_1 = \mathcal{O}_P((\log h^{-1})^{-2}\|\chi^{(2)}_{q_\tau^+}\|^2_{L^2})$.

To bound T_2, we estimate with (5.45) on $\Omega_{n,h}$

$$\begin{aligned}T_2 \lesssim&\eta_n^{-4}\int(u^2+1)|\mathcal{F}K(hu)|^2\big|\widetilde{\varphi}_{T,n}^{-1}(u)-\varphi_T^{-1}(u)\big|^2\mathrm{d}u\\ \leqslant&\eta_n^{-4}\int(u^2+1)|\mathcal{F}K(hu)|^2|\varphi_T(u)|^{-4}(|\Phi_n(u)|^2+n(u^2+u^4)^{-1}|\Phi_n(u)|^4)\mathrm{d}u\end{aligned}$$

Using the estimate (5.43), Φ_n is a centered normal random variable with variance smaller than $n^{-1}u^2(u^2+1)\|e^{-x}\rho\|^2_{L^2}$. Therefore,

$$\begin{aligned}T_2 &= \mathcal{O}_P\Big(n^{-1}\eta_n^{-4}\int u^2(u^2+1)^2|\mathcal{F}K(hu)|^2|\varphi_T(u)|^{-4})\mathrm{d}u\Big)\\ &= \mathcal{O}_P\Big(n^{-1}\eta_n^{-4}h^{-4}\sup_{|v|\leqslant 1/h}|\varphi_T(v)|^{-2}\|\chi^{(2)}_{q_\tau^+}\|^2_{L^2}\Big)\\ &= \mathcal{O}_P\Big((\log h^{-1})^{-2}\|\chi^{(2)}_{q_\tau^+}\|^2_{L^2}\Big).\end{aligned}$$

We conclude (5.55) for $k=2$. For $k=0,1$ similar calculations apply using

$$\begin{aligned}(\widetilde{\varphi}_{T,n}^{-1}-\varphi_T^{-1})' &= \frac{-\Phi_n^{(1)}+T\psi'\Phi_n}{\widetilde{\varphi}_{T,n}^2}+T\psi'(\varphi_T^{-1}-\widetilde{\varphi}_{T,n}^{-1}),\\ (\widetilde{\varphi}_{T,n}^{-1}-\varphi_T^{-1})'' &= \frac{-\Phi_n^{(2)}+T\psi''\Phi_n+T\psi'\Phi_n^{(1)}}{\widetilde{\varphi}_{T,n}^2}\\ &\quad-\frac{2}{\widetilde{\varphi}_{T,n}^2}\Big(1-\frac{\Phi_n}{\widetilde{\varphi}_{T,n}}\Big)\Big(T\psi'+\frac{\Phi_n^{(1)}}{\varphi_T}\Big)\big(-\Phi_n^{(1)}+T\psi'\Phi_n\big)\\ &\quad+T\big(\psi''(\varphi_T^{-1}-\widetilde{\varphi}_{T,n}^{-1})+\psi'(\varphi_T^{-1}-\widetilde{\varphi}_{T,n}^{-1})'\big).\end{aligned}$$ □

With these preparations at hand, we can proceed exactly as in the proof of Theorem 4.12 to see that $\widetilde{h}^{\pm}$ mimics the oracle bandwidth which balances the deterministic error Dh^{s+1} and the stochastic error $\widetilde{S}_{n,h}$ in the numerator of (5.53). Details are omitted.

A. Appendix: Function spaces

Let us define the L^p-Sobolev space for $p \in (0, \infty)$ and $m \in \mathbb{N}$

$$W_p^m(\mathbb{R}) := \Big\{ f \in L^p(\mathbb{R}) \Big| \sum_{k=0}^{m} \|f_X^{(k)}\|_{L^p} < \infty \Big\}$$

In particular, $W_p^0(\mathbb{R}) = L^p(\mathbb{R})$. Due to the Hilbert space structure, the case $p = 2$ is crucial. It can be described equivalently with the notation $\langle u \rangle = (1 + u^2)^{1/2}$ by, $\alpha \geqslant 0$,

$$H^\alpha(\mathbb{R}) := \Big\{ f \in L^2(\mathbb{R}) \Big| \|f\|_{H^\alpha}^2 := \int \langle u \rangle^{2\alpha} |\mathcal{F} f(u)|^2 \mathrm{d}u < \infty \Big\} \tag{A.1}$$

which we call Sobolev space, too. Obviously, $W_2^m(\mathbb{R}) = H^m(\mathbb{R})$. Also frequently used are the Hölder spaces. Denoting $\langle \alpha \rangle$ as the largest integer strictly smaller than α, we define for some function f and any possibly unbounded interval $I \subseteq \mathbb{R}$ the Hölder norm

$$\|f\|_{C^\alpha(I)} := \sum_{k=0}^{\langle \alpha \rangle} \|f^{(k)}\|_{L^\infty(I)} + \sup_{x,y \in I : x \neq y} \frac{|f^{\langle \alpha \rangle}(x) - f^{\langle \alpha \rangle}(y)|}{|x - y|^{\alpha - \langle \alpha \rangle}}. \tag{A.2}$$

Let $C^0(I)$ denote the space of all continuous and bounded functions on the interval I and

$$C^s(\mathbb{R}) := \bigcup_{R>0} C^s(\mathbb{R}, R) \quad \text{with} \quad C^\alpha(I, R) := \{ f \in C^0(I) | \|f\|_{C^\alpha(I)} < R \}, R > 0. \tag{A.3}$$

A unifying approach which contains all function spaces defined so far, is given by Besov spaces (Triebel, 2010, Sect. 2.3.1) which we will discuss in the sequel. Let $\mathscr{S}(\mathbb{R})$ be the Schwartz space of all rapidly decreasing infinitely differentiable functions with values in $\mathbb{C}$ and $\mathscr{S}'(\mathbb{R})$ its dual space, that is the space of all tempered distributions. Let $0 < \psi \in \mathscr{S}(\mathbb{R})$ with $\operatorname{supp} \psi \subseteq \{x | 1/2 \leqslant |x| \leqslant 2\}$ and $\psi(x) > 0$ if $\{x | 1/2 < |x| < 2\}$. Then define $\varphi_j(x) := \psi(2^{-j}x)(\sum_{k=-\infty}^{\infty} \psi(2^{-k}x))^{-1}, j = 1, 2, \dots,$ and $\varphi_0(x) := 1 - \sum_{j=1}^{\infty} \varphi_j(x)$ such that the sequence $\{\varphi_j\}_{j=0}^{\infty}$ is a smooth resolution of unity. In particular, $\mathcal{F}^{-1}[\varphi_j \mathcal{F} f]$ is an entire function for all $f \in \mathscr{S}'(\mathbb{R})$. For $s \in \mathbb{R}$ and $p, q \in (0, \infty]$ the Besov spaces are defined by

$$B_{p,q}^s(\mathbb{R}) := \Big\{ f \in \mathscr{S}'(\mathbb{R}) \Big| \|f\|_{B_{p,q}^s} := \Big(\sum_{j=0}^{\infty} 2^{sjq} \| \mathcal{F}^{-1}[\varphi_j \mathcal{F} f] \|_{L^p}^q \Big)^{1/q} < \infty \Big\}.$$

We omit the dependence of $\|\bullet\|_{B_{p,q}^s}$ to ψ since any function with the above properties defines an equivalent norm. Setting the Besov spaces in relation to the more elementary function spaces, we first note that the Schwartz functions $\mathscr{S}(\mathbb{R})$ are dense in every Besov space $B_{p,q}^s(\mathbb{R})$ with $p, q < \infty$ and $H^\alpha(\mathbb{R}) = B_{2,2}^\alpha(\mathbb{R})$ as well as $C^\alpha(\mathbb{R}) = B_{\infty,\infty}^\alpha(\mathbb{R})$, where the latter holds only if α is not an integer (Triebel, 2010, Thms. 2.3.3 and 2.5.7). Frequently used are the following continuous embeddings which can be found in (Triebel, 2010, Sect. 2.5.7, Thms. 2.3.2(1), 2.7.1): For $p \geqslant 1, m \in \mathbb{Z}$

$$B^m_{p,1}(\mathbb{R}) \subseteq W^m_p(\mathbb{R}) \subseteq B^m_{p,\infty}(\mathbb{R}) \quad \text{and} \quad B^0_{\infty,1}(\mathbb{R}) \subseteq L^\infty(\mathbb{R}) \subseteq B^0_{\infty,\infty}(\mathbb{R}) \tag{A.4}$$

and for $s \geqslant 0$

$$B^s_{\infty,1}(\mathbb{R}) \subseteq C^s(\mathbb{R}) \subseteq B^s_{\infty,\infty}(\mathbb{R}). \tag{A.5}$$

Furthermore, for $0 < p_0 \leqslant p_1 \leqslant \infty, q \geqslant 0$ and $-\infty < s_1 \leqslant s_0 < \infty$

$$B^{s_0}_{p_0,q}(\mathbb{R}) \subseteq B^{s_1}_{p_1,q}(\mathbb{R}) \quad \text{if} \quad s_0 - \frac{1}{p_0} \geqslant s_1 - \frac{1}{p_1} \tag{A.6}$$

and for $0 < p, q_0, q_1 \leqslant \infty$ and $-\infty < s_1 < s_0 < \infty$

$$B^{s_0}_{p,q_0}(\mathbb{R}) \subseteq B^{s_1}_{p,q_1}(\mathbb{R}). \tag{A.7}$$

Another important relation is the pointwise multiplier property of Besov spaces (Triebel, 2010, (24) on p. 143) that is

$$\|fg\|_{B^s_{p,q}} \lesssim \|f\|_{B^s_{\infty,q}} \|g\|_{B^s_{p,q}} \tag{A.8}$$

for $s > 0$, $1 \leqslant p \leqslant \infty$ and $0 < q \leqslant \infty$.

The Besov norm of a convolution can be bounded by Lemma 7 (i) by Qui (1981). Let $1 \leqslant p, q, r, s \leqslant \infty$, $-\infty < \alpha, \beta < \infty$, $0 \leqslant 1/u = 1/p + 1/r - 1 \leqslant 1$, $0 \leqslant 1/v = 1/q + 1/s \leqslant 1$. For $f \in B^\alpha_{p,q}(\mathbb{R})$ and $g \in B^\beta_{r,s}(\mathbb{R})$

$$\|f * g\|_{B^{\alpha+\beta}_{u,v}} \lesssim \|f\|_{B^\alpha_{p,q}} \|g\|_{B^\beta_{r,s}}. \tag{A.9}$$

The space of finite signed Borel measures on the real line will be denoted by $\mathcal{M}(\mathbb{R})$. For any $\mu \in \mathcal{M}(\mathbb{R})$ there are two positive finite measures μ^+, μ^- such that $\mu(A) = \mu^+(A) - \mu^-(A)$ for any $A \in \mathscr{B}(\mathbb{R})$. Using this so-called Jordan-decomposition, the total variation norm on $\mathcal{M}(\mathbb{R})$ is defined by

$$\|\mu\|_{TV} := \mu^+(\mathbb{R}) + \mu^-(\mathbb{R}).$$

If a function $f \colon \mathbb{R} \to \mathbb{R}$ is locally integrable, it defines a distribution $T_f(\psi) = \int \psi f$ on the test function space $\mathcal{D}(\mathbb{R})$ of infinitely smooth functions with bounded support. The distributional derivative of T_f is denoted by $Df := \rho$. The space of functions of bounded variation is then defined by

$$BV(\mathbb{R}) := \{f \colon \mathbb{R} \to \mathbb{R} \text{ locally integrable}, Df \in \mathcal{M}(\mathbb{R})\}$$

with bounded variation norm $\|f\|_{BV} := \|Df\|_{TV}$ for $f \in BV(\mathbb{R})$. Slightly abusing notation by identifying $f \in BV(\mathbb{R})$ with its equivalence class with respect to equality almost everywhere, the relation of $BV(\mathbb{R})$ to Besov spaces is given by (Giné and Nickl, 2008, Lem. 8):

$$BV(\mathbb{R}) \cap L^1(\mathbb{R}) \subseteq B^1_{1,\infty}(\mathbb{R}). \tag{A.10}$$

Bibliography

Aït-Sahalia, Y. and Jacod, J. (2009). Estimating the degree of activity of jumps in high frequency data. *Ann. Statist.*, 37(5A):2202 – 2244.

Aït-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *J. Econ. Lit.*, 50(4):1007–50.

Asmussen, S., Avram, F., and Pistorius, M. R. (2004). Russian and American put options under exponential phase-type Lévy models. *Stochastic Proc. Appl.*, 109(1):79 – 111.

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse gaussian type. *Finance Stoch.*, 2:41–68.

Basawa, I. V. and Brockwell, P. J. (1982). Nonparametric estimation for nondecreasing Lévy processes. *J. Roy. Statist. Soc. Ser. B Stat. Methodol.*, 44(2):262–269.

Bauer, F. and Reiß, M. (2008). Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Problems*, 24(5):1–16.

Belomestny, D. (2010). Spectral estimation of the fractional order of a Lévy process. *Ann. Statist.*, 38(1):317–351.

Belomestny, D. (2011). Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.*, 39(4):2205–2242.

Belomestny, D. and Reiß, M. (2006a). Spectral calibration of exponential Lévy models. *Finance Stoch.*, 10(4):449–474.

Belomestny, D. and Reiß, M. (2006b). Spectral calibration of exponential Lévy models [2]. SFB 649 Discussion Paper 2006-035, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. Available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2006-035.pdf.

Belomestny, D. and Schoenmakers, J. (2011). A jump-diffusion libor model and its robust calibration. *Quant. Finance*, 11(4):529–546.

Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

Bickl, P. J. and Ritov, Y. (2003). Nonparametric estimators which can be ”plugged-in”. *Ann. Statist.*, 31(4):1033–1053.

Bissantz, N., Dümbgen, L., Holzmann, H., and Munk, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(3):483–506.

Bissantz, N., Hohage, T., and Munk, A. (2004). Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. *Inverse Problems*, 20(6):1773–1789.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81(3):637–654.

Blumenthal, R. M. and Getoor, R. K. (1961). Sample functions of stochastic processes with stationary independent increments. *J. Math. Mech.*, 10:493–516.

Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6):2384–2398.

Buchmann, B. and Grübel, R. (2003). Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.*, 31(4):1054–1074.

Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, 15(1):69–98.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Bus.*, 75(2):305–332.

Carr, P. and Madan, D. B. (1999). Option valuation using the fast Fourier transform. *J. Comput. Finance*, 2:61–73.

Carroll, R. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83(404):1184–1186.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*, volume 105. Chapman & Hall/CRC, Boca Raton, FL.

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 19.

Cavalier, L., Golubev, G. K., Picard, D., and Tsybakov, A. B. (2002). Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3):843–874. Dedicated to the memory of Lucien Le Cam.

Comte, F. and Genon-Catalot, V. (2011). Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.*, 39(2):803–837.

Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(4):601–627.

Cont, R. and Tankov, P. (2004a). *Financial modelling with jump processes*. Chapman & Hall / CRC Press, Boca Raton, FL.

Cont, R. and Tankov, P. (2004b). Non-parametric calibration of jump-diffusion option pricing models. *J. Comput. Finance*, 7(3):1–49.

Cont, R. and Voltchkova, E. (2005). Integro-differential equations for option prices in exponential Lévy models. *Finance Stoch.*, 9:299–325.

Da Prato, G. and Zabczyk, J. (1992). *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

Dattner, I., Goldenshluger, A., and Juditsky, A. (2011). On deconvolution of distribution functions. *Ann. Statist.*, 39(5):2477–2501.

Dattner, I. and Reiser, B. (2013). Estimation of distribution functions in measurement error models. *J. Statist. Plann. Inference*, 143(3):479–493.

Dattner, I., Reiß, M., and Trabs, M. (2013). Adaptive estimation of quantiles in deconvolution with unknown error distribution. *arXiv*, 1303.1698.

Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, 36(2):665–685.

Dudley, R. M. (1992). Fréchet differentiability, p-variation and uniform Donsker classes. *Ann. Probab.*, 20(4):1968–1982.

Eberlein, E., Keller, U., and Prause, K. (1998). New insights into smile, mispricing and value at risk: The hyperbolic model. *J. Bus.*, 71:371–406.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 97–110.

Fan, J. (1991b). On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *Ann. Statist.*, 19(3):1257–1272.

Figueroa-López, J. (2011). Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli*, 17(2):643–670.

Figueroa-López, J. E. and Houdré, C. (2006). Risk bounds for the non-parametric estimation of Lévy processes. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 96–116. Inst. Math. Statist., Beachwood, OH.

Föllmer, H. and Leukert, P. (1999). Quantile hedging. *Finance Stoch.*, 3(3):251–273.

Frese, E. M., Fick, A., and Sadowsky, H. S. (2011). Blood pressure measurement guidelines for physical therapists. *Cardiopulmonary physical therapy journal*, 22(2):5–12.

Giné, E. and Nickl, R. (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields*, 141(3-4):333–387.

Girardi, M. and Weis, L. (2003). Operator-valued Fourier multiplier theorems on Besov spaces. *Math. Nachr.*, 251(1):34–51.

Goldenshluger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.*, 6(2):135–170.

Goldenshluger, A. and Pereverzev, S. V. (2000). Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Related Fields*, 118(2):169–186.

Goldenshluger, A. and Pereverzev, S. V. (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli*, 9(5):783–807.

Grama, I. and Nussbaum, M. (2002). Asymptotic equivalence for nonparametric regression. *Math. Methods Statist.*, 11(1):1–36.

Gugushvili, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.*, 21(3):321–343.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 14:323–330.

Hall, P. and Lahiri, S. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist.*, 36(5):2110–2134.

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press, Cambridge.

Hartman, P. and Wintner, A. (1942). On the infinitesimal generators of integral convolutions. *Amer. J. Math.*, 64:273–298.

Ivanov, R. V. (2007). On the Pricing of American Options in Exponential Lévy Markets. *J. Appl. Probab.*, 44(2):409–419.

Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.*, 37(5A):2301–2323.

Johannes, J. and Schwarz, M. (2013). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli*, 19(5A):1576–1611.

Kallenberg, O. (1981). Splitting at backward times in regenerative sets. *Ann. Probab.*, 9(5):781–799.

Kannel, W. et al. (1995). Framingham study insights into hypertensive risk of cardiovascular disease. *Hypertension research: official journal of the Japanese Society of Hypertension*, 18(3):181–196.

Kappus, J. (2012). *Nonparametric adaptive estimation for discretely observed Lévy processes.* PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.

Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Process. Appl.*, 124(1):730 – 758.

Kappus, J. and Reiß, M. (2010). Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl.*, 64(3):314–328.

Khoujmane, A., Ruymgaart, F., and Shubov, M. (2007). The convolution theorem for estimating linear functionals in indirect nonparametric regression models. *J. Statist. Plann. Inference*, 137(3):811–820.

Klaassen, C. A. J., Lee, E.-J., and Ruymgaart, F. H. (2001). On efficiency of indirect estimation of nonparametric regression functions. In *Algebraic methods in statistics and probability (Notre Dame, IN, 2000)*, volume 287 of *Contemp. Math.*, pages 173–184. Amer. Math. Soc., Providence, RI.

Klaassen, C. A. J., Lee, E.-J., and Ruymgaart, F. H. (2005). Asymptotically efficient estimation of linear functionals in inverse regression models. *J. Nonparametr. Stat.*, 17(7):819–831.

Klenke, A. (2007). *Probability Theory: A Comprehensive Course (Universitext)*. Springer, London, 1st edition.

Knopova, V. and Schilling, R. L. (2013). A note on the existence of transition probability densities of Lévy processes. *Forum Math.*, 25(1):125–149.

Kuo, H. H. (1975). *Gaussian measures in Banach spaces.* Lecture Notes in Mathematics, Vol. 463. Springer-Verlag, Berlin.

Le Cam, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics*, pages 245–261, Berkeley, Calif. Univ. California Press.

Lepski, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470.

Liese, F. (1987). Estimates of Hellinger integrals of infinitely divisible distributions. *Kybernetika*, 23(3):227–238.

Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.*, 39(1):201–231.

Madan, D. B. and Seneta, E. (1990). The variance gamma (VG) model for share market returns. *J. Bus.*, 63:511–524.

Massart, P. (2007). *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics.* Springer, Berlin.

Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canad. J. Stat.*, 32(4):439–449.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Lecture notes in statistics. Springer, Berlin Heidelberg.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *J. Finan. Econ.*, 3(1-2):125 – 144.

Natterer, F. (1984). Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Anal.*, 18(1-2):29–37.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Stat.*, 7(4):307–330.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.*, 98(10):1955–1968.

Neumann, M. H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248.

Nickl, R. and Reiß, M. (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal.*, 263(10):3306–3332.

Nickl, R., Söhl, J., Reiß, M., and Trabs, M. (2013). High-frequency Donsker theorems for Lévy measures. arXiv:1310.2523.

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and gaussian white noise. *Ann. Statist.*, 24(6):2399–2430.

Orey, S. (1968). On continuity properties of infinitely divisible distribution functions. *Ann. Math. Statist.*, 39:936–937.

Qui, B. H. (1981). Bernstein's theorem and translation invariant operators. *Hiroshima Math. J.*, 11(1):81–96.

Reiß, M. (2013). Testing the characteristics of a Lévy process. *Stochastic Process. Appl.*, 123(7):2808–2828. Special Issue International Year of Statistics.

Rosenthal, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.*, 8:273–303.

Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.

Schmidt–Hieber, J., Munk, A., and Dümbgen, L. (2013). Multiscale Methods for Shape Constraints in Deconvolution: Confidence Statements for Qualitative Features. *Annals of Statistics*, 41(3):1299–1328.

Söhl, J. (2010). Polar sets for anisotropic gaussian random fields. *Statist. Probab. Lett.*, 80(9-10):840 – 847.

Söhl, J. (2014). Confidence sets in nonparametric calibration of exponential Lévy models. *Finance Stoch.*, To appear. arXiv:1202.6611.

Söhl, J. and Trabs, M. (2012a). Documentation: Option calibration of exponential Lévy models in R. Technical report, Humboldt-Universität zu Berlin. Available at http://www.math.hu-berlin.de/˜trabs.

Söhl, J. and Trabs, M. (2012b). A uniform central limit theorem and efficiency for deconvolution estimators. *Electron. J. Statist.*, 6:2486–2518.

Söhl, J. and Trabs, M. (2014). Option calibration of exponential Lévy models: Confidence intervals and empirical results. *J. Comput. Finance*, To appear. arXiv 1202.5983.

Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498.

Stirnemann, J. J., Comte, F., and Samson, A. (2012). Density estimation of a biomedical variable subject to measurement error using an auxiliary set of replicate observations. *Statistics in Medicine*.

Trabs, M. (2013). Information bounds for inverse problems with application to deconvolution and Lévy models. arXiv:1307.6610.

Trabs, M. (2014). Calibration of self-decomposable Lévy models. *Bernoulli*, 20(1):109–140.

Triebel, H. (2010). *Theory of Function Spaces.* Birkhäuser Verlag, Basel. Reprint of the 1983 Edition.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

van der Vaart, A. W. (1988). *Statistical estimation in large parameter spaces*, volume 44 of *CWI Tract*. Stichting Mathematisch Centrum Centrum voor Wiskunde en Informatica, Amsterdam.

van der Vaart, A. W. (1991). On differentiable functionals. *Ann. Statist.*, 19(1):178–204.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer, New York.

van Es, B. and Uh, H.-W. (2005). Asymptotic normality of kernel-type deconvolution estimators. *Scand. J. Statist.*, 32(3):467–483.

van Rooij, A. C. M., Ruymgaart, F. H., and van Zwet, W. R. (1999). Asymptotic efficiency of inverse estimators. *Teor. Veroyatnost. i Primenen.*, 44(4):826–844. translation in *Theory Probab. Appl.* 44(4), 722-738.

Witting, H. (1985). *Mathematische Statistik. I.* B. G. Teubner, Stuttgart. Parametrische Verfahren bei festem Stichprobenumfang. [Parametric methods for fixed sample size].

List of Figures

List of Tables

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 27.01.2014 Mathias Trabs